

# *Modelling the structure of a multimodal artefact*



*Tuomo Hiippala*

FACULTY OF ARTS  
UNIVERSITY OF HELSINKI

# Modelling the structure of a multimodal artefact

Tuomo Hiippala

Department of Modern Languages  
University of Helsinki

ACADEMIC DISSERTATION

To be publicly discussed, by due permission of the Faculty of Arts at the University of Helsinki in lecture room 5, University Main Building, on the 14<sup>th</sup> of December, 2013 at 10 o'clock.

© Tuomo Hiippala 2013

Typeset in Computer Modern using L<sup>A</sup>T<sub>E</sub>X

Cover design by Valpuri Liimatainen

ISBN 978-952-10-9426-2 (paperback)

ISBN 978-952-10-9427-9 (PDF)

<http://ethesis.helsinki.fi>

Helsinki 2013

Unigrafia

# Abstract

This dissertation studied the structure of multimodal artefacts, or how language, image and other semiotic modes combine and interact in documents. This places the study within the emerging field of multimodal research, which uses linguistic methods to study the interaction of multiple semiotic modes.

Despite the growing amount of multimodal research, the structure of multimodal artefacts has not received the attention it warrants. Previous studies have been either very detailed or exceedingly abstract, leaving a significant gap between data and theory, which this dissertation attempted to bridge. To do so, the dissertation adopted a data-driven approach to multimodal analysis, addressing the structure of multimodal artefacts, the factors that shape the artefact structure, and the role of structure in the recognition and interpretation of the artefacts.

The data consisted of tourist brochures produced by the city of Helsinki between 1967 and 2008, which allowed a longitudinal perspective to their multimodal structure. A total of 58 double-pages were annotated for their content, visual appearance, layout and rhetorical organisation, and compiled into an XML-based multimodal corpus. To study the corpus, the dissertation developed visualisation methods that combined information from multiple analytical layers of the corpus to represent the multimodal structures in the data.

The study revealed the functional motivation behind the structure of the tourist brochures, identifying patterns in their hierarchical and rhetorical organisation, which were used to fulfil specific communicative tasks. The configuration of these patterns, in turn, signalled how the brochure was to be interpreted. The results also showed that after the year 1985, which marked the introduction of desktop publishing software, the organising principles of the tourist brochures have shifted towards a more fragmented and non-linear structure.

# Acknowledgements

First of all, I would like to thank my supervisor, professor Eija Ventola, whom I first met back in 2005 when I was still an undergraduate student. Looking back at the previous eight years, I could not have hoped for a better supervisor for my master's thesis and doctoral dissertation. Not only did Eija introduce me to the entire field of study, but she has been encouraging and inspirational throughout these years: I do not think I should thank her only for supervising my work, but for letting me take my time to develop as a researcher.

Eija once told me that the most important thing a supervisor can do is to introduce the student to the international research community. Not surprisingly, Eija's international contacts proved invaluable during my doctoral studies. In 2008, Eija introduced me to professor John Bateman — one of the two pre-examiners of this dissertation — at the 4<sup>th</sup> International Conference on Multimodality in Singapore, organised by my other pre-examiner and opponent, professor Kay O'Halloran. I am grateful to both of them for their contribution to this dissertation.

With John's support and a grant provided by the Deutscher Akademischer Austauschdienst, I was able to spend six months at the University of Bremen in early 2009, which greatly helped me to focus my work. Ritva and Volker Saegert were kind enough to host me for the first couple of weeks, although we had not seen each other for 20 years — Vielen Dank! During the time I also shared an office with Chiaoi Tseng, whom I wish to thank for her company and for her friendly advice to a beginner. I also wish to thank Fabian Töpel and Nawid Ketabi for accommodating me during my many weekends in Köln. The same applies to Kariina Lehtovaara-Skrifvars and Markus Skrifvars, who let me stay as their guest for an entire month in 2010, when I attended a conference in Sydney, Australia. Finally, Mika Meskanen deserves thanks for hosting me countless times in both Berlin and London.

During all of this time, my closest colleague in Helsinki has been Ivan Berazhny. He has not only become a good friend, but he is also one of the most helpful people I have met. Ivan has always been willing to comment, proofread and offer his advice on my work. Thank you, Ivan, and everyone else at the Department of Modern Languages as well!

When it comes to working space, the post-graduate student today is like a piece in a game of Tetris: you get moved around a lot. So far, I have enjoyed the company of Jari Käkälä, Elina Liikanen, Iris Rennie and Hanna Lantto — thank you all. Most recently, I have shared an office with Vappu Kannas. Although we deal with completely different topics, there is always something to talk about and that has made working fun. Thank you, Vappu!

Due to the numerous travel grants from the Department of Modern Languages and the Chancellor of the University of Helsinki, I have been able to present my work at conferences both home and abroad, where I have met researchers with similar interests: I am grateful for these opportunities. Martin Thomas, Janina Wildfeuer, Morten Boeriis, Christian Johannessen, William Feng, Gunhild Kvåle, Victor Lim Fei, Bradley Smith, Kaela Zhang and many others: thank you all for the inspiring discussions.

My family has been supportive of my work at all times: I thank my parents — Anita and Seppo — for their patience and help. I also wish to thank my sisters Saara and Kaisa, and my brother Anssi. To all of my friends: I am lucky, because there are way too many of you to be listed here — thank you all! Valpuri: thank you for your constant support and understanding, and for the cover of this dissertation.

Finally, I wish to extend my thanks to the foundations of Emil Aaltonen and Ella & Georg Ehrnrooth for allowing me to concentrate on finishing this dissertation.

Helsinki, October 2013

# Permissions

All excerpts from the tourist brochures included in this dissertation are used with permission of Helsinki City Tourist & Convention Bureau.

Figure 3.5, originally published in W. Bowcher (ed.), *Multimodal Texts from Around the World: Cultural and Linguistic Insights*, 2012, Palgrave MacMillan, reproduced with permission of Palgrave Macmillan.

# Contents

<b>List of Figures</b>	<b>xii</b>
<b>List of Tables</b>	<b>xiv</b>
<b>1 Introduction</b>	<b>1</b>
1.1 Why use linguistic methods for multimodal research? . . . . .	2
1.2 Why do empirical research on multimodality? . . . . .	3
1.3 Why model the structure of a multimodal artefact? . . . . .	5
1.4 Why study the tourist brochures? . . . . .	6
1.4.1 The tourist brochure as a multimodal artefact . . . . .	7
1.4.2 Commercial factors and applications . . . . .	7
1.4.3 Access to research material . . . . .	8
1.5 Research questions and the structure of the dissertation . . . . .	9
1.5.1 Research questions . . . . .	9
1.5.2 Dissertation structure . . . . .	9
<b>2 Multimodality as a field of study</b>	<b>11</b>
2.1 Origins and development . . . . .	11
2.1.1 Origins . . . . .	11
2.1.2 Influences and parallel research . . . . .	12
2.1.3 Current challenges . . . . .	14
2.2 Influence of systemic-functional linguistics and social semiotics . . . . .	18
2.2.1 Overview of the underlying theories . . . . .	19
2.2.2 Implications to multimodality . . . . .	21
2.3 Key theoretical concepts . . . . .	22
2.3.1 Mode . . . . .	23
2.3.1.1 Mode in linguistics . . . . .	23
2.3.1.2 Mode in multimodal research . . . . .	25
2.3.1.3 Modal interrelations . . . . .	25
2.3.1.4 Mode in social semiotics . . . . .	28
2.3.2 Rank scale . . . . .	30



2.3.3	Metafunctions . . . . .	31
2.4	Social semiotic analyses of print media . . . . .	34
2.4.1	Early work . . . . .	34
2.4.2	<i>Visual grammar</i> . . . . .	35
2.4.2.1	Transitivity . . . . .	37
2.4.2.2	Modality . . . . .	38
2.4.2.3	Composition . . . . .	40
2.5	Layout in print media . . . . .	42
2.5.1	Intersemiotic complementarity . . . . .	43
2.5.2	Analysing print media using SF-MDA . . . . .	45
2.5.3	Generic structures . . . . .	47
2.5.4	Covers as contact texts . . . . .	48
2.6	Concluding remarks . . . . .	48
<b>3</b>	<b>An empirical approach to multimodality</b>	<b>50</b>
3.1	The <i>Genre and Multimodality</i> (GeM) model . . . . .	50
3.2	The analytical layers of the GeM model . . . . .	53
3.2.1	Base layer . . . . .	53
3.2.2	Layout layer . . . . .	55
3.2.2.1	Layout structure . . . . .	55
3.2.2.2	Area model . . . . .	58
3.2.2.3	Realisation information . . . . .	61
3.2.3	Rhetorical layer . . . . .	62
3.2.3.1	Classical RST . . . . .	64
3.2.3.2	Extended GeM RST . . . . .	65
3.2.4	Navigation layer . . . . .	67
3.3	The concept of a semiotic mode . . . . .	68
3.3.1	Material substrate . . . . .	69
3.3.2	Semiotic resources . . . . .	70
3.3.3	Discourse semantics . . . . .	73
3.4	Semiotic modes in the tourist brochures . . . . .	74
3.4.1	<i>Text-flow</i> . . . . .	74
3.4.2	<i>Image-flow</i> . . . . .	76
3.5	<i>Page-flow</i> in the tourist brochures . . . . .	77
3.5.1	The page as a material substrate . . . . .	78
3.5.2	The semiotic resources on a page . . . . .	79
3.5.3	Discourse semantics in the tourist brochures . . . . .	80
3.6	Concluding remarks . . . . .	82

<b>4</b>	<b>Genre</b>	<b>84</b>
4.1	An outlook on deploying genre . . . . .	85
4.1.1	On the notion of genre . . . . .	85
4.1.2	The role of criteria in defining genre . . . . .	87
4.2	The study of rhetoric: genre as social action . . . . .	88
4.3	Linguistic studies of genre in ESP and SFL . . . . .	90
4.3.1	Staging the tourist brochures . . . . .	92
4.3.2	Genre dynamics . . . . .	93
4.3.2.1	Semogenic processes . . . . .	94
4.3.2.2	Genre agnation . . . . .	96
4.4	Genre in document theory and information design . . . . .	98
4.4.1	Genre as an expectation-generating device . . . . .	100
4.4.2	Layout as a memory tool . . . . .	100
4.4.3	Document design as a pattern language . . . . .	102
4.5	A framework for modelling the structure of a multimodal artefact . . . . .	104
<b>5</b>	<b>Data and methods</b>	<b>105</b>
5.1	Data . . . . .	105
5.1.1	Sources and digitalisation . . . . .	105
5.1.2	Representativeness of the annotated corpus . . . . .	106
5.1.3	The two data sets . . . . .	106
5.1.3.1	The entire data set . . . . .	107
5.1.3.2	The annotated corpus . . . . .	108
5.2	Methods: applying the GeM model . . . . .	109
5.2.1	Data segmentation: base layer . . . . .	110
5.2.2	Composition and appearance: layout layer . . . . .	111
5.2.2.1	Area model . . . . .	112
5.2.2.2	Layout structure . . . . .	114
5.2.2.3	Realisation information . . . . .	115
5.2.3	Presenting the content: rhetorical layer . . . . .	116
5.2.4	Guiding the reader: navigation layer . . . . .	117
5.3	Building the corpus . . . . .	118
5.3.1	File names . . . . .	119
5.3.2	Document Type Definitions (DTD) . . . . .	119
5.4	Verifying the corpus . . . . .	120
5.5	Exploring the corpus . . . . .	122
5.5.1	Query languages . . . . .	122
5.5.2	Visualisation techniques . . . . .	123

<b>6</b>	<b>The medium and its characteristics</b>	<b>127</b>
6.1	Why the notion of a medium matters . . . . .	128
6.1.1	The brochure as a medium . . . . .	129
6.1.2	Brochures and leaflets . . . . .	130
6.2	The relationship between a medium and advertising . . . . .	133
6.2.1	A multimodal perspective to advertising . . . . .	133
6.2.2	Advertising in the tourist brochures . . . . .	136
6.3	Aspects of production . . . . .	138
6.3.1	The ‘visual turn’ in the tourist brochures . . . . .	139
6.3.2	The impact of technological development . . . . .	143
6.4	Summary . . . . .	146
<b>7</b>	<b>The content and its structure</b>	<b>148</b>
7.1	Representing the destination using text-flow . . . . .	149
7.1.1	Joint: the use of tables . . . . .	151
7.1.2	Elaboration in descriptive texts . . . . .	153
7.1.3	Enablement: supporting the descriptions . . . . .	156
7.1.4	Sequences: history and guidance . . . . .	157
7.2	Graphic elements in the tourist brochures . . . . .	159
7.2.1	Photographs in image-text-complexes . . . . .	160
7.2.2	Rhetorically weak: conceptual photographs . . . . .	163
7.3	Illustrations and maps . . . . .	167
7.3.1	Illustrations as hand-drawn representations . . . . .	167
7.3.2	Illustrations and identification . . . . .	169
7.3.3	Maps . . . . .	171
7.4	Summary . . . . .	173
<b>8</b>	<b>The page and its interpretation</b>	<b>175</b>
8.1	The step from text-flow to page-flow . . . . .	176
8.2	Investigating page-flow . . . . .	179
8.2.1	Interrogating the layout structure . . . . .	180
8.2.2	Layout structures in the annotated corpus . . . . .	186
8.3	Formalising the discourse semantics . . . . .	191
8.4	Text-flow and page-flow in the annotated corpus . . . . .	196
8.4.1	Detecting the active semiotic mode . . . . .	197
8.4.2	A closer look at page-flow . . . . .	200
8.5	Semiotic modes and semogenic processes . . . . .	201
8.6	Summary . . . . .	204

<b>9</b>	<b>Discussion</b>	<b>206</b>
9.1	Main findings . . . . .	206
9.1.1	Which factors affect the structure of a multimodal artefact?	206
9.1.2	What kinds of patterns may be identified in the multimodal structure? . . . . .	207
9.1.3	How does the structure of the multimodal artefacts change over time? . . . . .	208
9.2	Relevance of the findings . . . . .	209
<b>10</b>	<b>Conclusion</b>	<b>212</b>
10.1	Summary . . . . .	212
10.2	Evaluation of the study . . . . .	213
10.3	Implications . . . . .	215
10.3.1	Contributions to multimodal research . . . . .	215
10.3.2	Pragmatic advice on tourist brochure design . . . . .	216
10.4	Avenues of further research . . . . .	218
	<b>Appendix A The entire data set</b>	<b>220</b>
	<b>Appendix B The rhetorical relations used in GeM RST</b>	<b>224</b>
	<b>Bibliography</b>	<b>227</b>

# List of Figures

2.1	A system network . . . . .	20
2.2	Linguistic mode in Halliday (1978) . . . . .	24
2.3	Network of modes, sub-modes and features in printed media . . . . .	27
2.4	A double-page from <i>Holiday in Helsinki</i> (1985) . . . . .	36
2.5	Aspects of visual grammar in the analysis of print media . . . . .	40
2.6	SF-MDA framework for printed texts: language and visual imagery . . . . .	46
3.1	The evolution of multimodal artefacts and genres . . . . .	52
3.2	An example of a hierarchical layout structure . . . . .	56
3.3	<i>Helsinki's Four Tourist Islands</i> (1988) with a resolution reduction from 300 DPI to 5 DPI . . . . .	57
3.4	Three GeM area models . . . . .	59
3.5	Navigation structure in a tourist brochure . . . . .	67
3.6	The three strata of a semiotic mode . . . . .	69
3.7	Paradigmatic and syntagmatic choices in language . . . . .	71
3.8	Paradigmatic and syntagmatic choices in photography . . . . .	72
3.9	Three semiotic modes in printed multimodal artefacts . . . . .	75
3.10	<i>Bohemian Nordic Oddity</i> (2006) . . . . .	76
3.11	An extract from <i>Suomenlinna Seafortress in Helsinki</i> (1988) . . . . .	81
4.1	Genre in SFL . . . . .	91
4.2	A schematic representation of content in the tourist brochures . . . . .	93
5.1	The cross-referencing of GeM annotation . . . . .	110
5.2	Area model of <i>Helsinki's Four Tourist Islands</i> (1972) . . . . .	113
5.3	An orphan span in the annotation of <i>Welcome to Helsinki</i> (1998) . . . . .	121
5.4	A looping relation in the annotation of <i>Welcome to Helsinki</i> (1998) . . . . .	121
5.5	A key to the used visualisations . . . . .	125
6.1	The number of folds and content pages in brochures and leaflets . . . . .	131
6.2	The distribution of advertisements in the tourist brochures . . . . .	137
6.3	Classified advertisements in <i>Weekend Delight in Helsinki</i> (2006) . . . . .	138

6.4	Graphic elements in the annotated corpus . . . . .	140
6.5	Two examples of technological development . . . . .	144
6.6	Maps in <i>Helsinki's Four Tourist Islands</i> . . . . .	145
7.1	The layout structure in <i>MEH</i> 1967 . . . . .	151
7.2	Examples of ELABORATION in <i>WEH</i> 1998 . . . . .	154
7.3	Enabling the tourist to perform activities in <i>HFT</i> 1972 . . . . .	156
7.4	Enabling the tourist to perform activities in <i>BNO</i> 2006 . . . . .	157
7.5	Telling the history of Suomenlinna using a SEQUENCE in <i>HFT</i> 1967 . . . . .	158
7.6	Guiding the reader using a SEQUENCE in <i>SHF</i> 1987 . . . . .	158
7.7	An image-text complex with a RESTATEMENT relation in <i>TGH</i> 1980 . . . . .	161
7.8	An image-text complex in <i>SSH</i> 1986 . . . . .	162
7.9	Conceptual photographs in <i>HFT</i> 1984 . . . . .	164
7.10	An illustration in an image-text-complex in <i>HDB</i> 1969 . . . . .	167
7.11	Examples of illustrations . . . . .	168
7.12	An illustration accompanying a header in <i>GNO</i> 2008 . . . . .	168
7.13	An illustration in an IDENTIFICATION relation in <i>HDB</i> 1972 . . . . .	169
7.14	An illustration functioning as a navigation structure in <i>HFT</i> 1972 . . . . .	170
7.15	Navigation structure in <i>WTH</i> 1983 . . . . .	172
8.1	<i>Welcome To Helsinki</i> (1998) . . . . .	177
8.2	An area model of <i>WEH</i> 1998 . . . . .	178
8.3	The layout and rhetorical structures in <i>WEH</i> 1998 . . . . .	178
8.4	The content side in <i>SSH</i> 1986 . . . . .	181
8.5	The layout structure of the fourth column in Figure 8.4 . . . . .	183
8.6	The cover side in <i>SSH</i> 1986 . . . . .	184
8.7	The layout structure of side one in <i>SSH</i> 1986 . . . . .	185
8.8	A double-page in <i>Helsinki Visitors Guide</i> (2008) . . . . .	187
8.9	The layout structure and a part of the RST structure in <i>HVG</i> 2008 . . . . .	189
8.10	Area-model of a single page in <i>HVG</i> 2008 . . . . .	193
8.11	Back-and-forth mappings of discourse semantics in <i>HVG</i> 2008 . . . . .	194
8.12	The distribution of semiotic modes in the annotated corpus . . . . .	198
9.1	Situating the model of a multimodal artefact . . . . .	210

# List of Tables

3.1	A list of base units . . . . .	54
4.1	Framing semiotic change . . . . .	94
4.2	Aginate genres among secondary school history genres . . . . .	96
5.1	Situational parameters listed as hierarchical sampling strata . . . . .	107
5.2	The annotated corpus . . . . .	108
5.3	Base layer elements and identifiers . . . . .	110
5.4	Layout layer elements and identifiers . . . . .	112
5.5	Entry types for the <i>type</i> attribute in the <i>graphics</i> element . . . . .	116
5.6	Rhetorical layer elements and identifiers . . . . .	117
5.7	Navigation layer identifiers . . . . .	118
6.1	The number of content pages in brochures and leaflets . . . . .	132
7.1	Rhetorical relations in the annotated corpus . . . . .	150
7.2	The distribution of graphic elements in the data . . . . .	160
7.3	Photographs in the rhetorical structure . . . . .	165
8.1	Layout chunks and layout leafs in <i>SSH</i> 1986 . . . . .	186
8.2	Pages with the highest number of layout chunks . . . . .	188
8.3	Page-flow and text-flow in the annotated corpus . . . . .	202
8.4	The degree of visuality in the annotated corpus . . . . .	203
8.5	The semiotic modes in the brochures and leaflets . . . . .	203
A.1	The entire data set . . . . .	221
B.1	List of asymmetric RST relations in the GeM RNG DTD . . . . .	224
B.2	List of symmetric RST relations in the GeM RNG DTD . . . . .	226
B.3	List of subnuclear RST relations in the GeM RNG DTD . . . . .	226

# Chapter 1

## Introduction

This dissertation studies the structure of documents, focusing on a particular document type — the tourist brochure. Like many other contemporary documents, the tourist brochures often communicate with the reader using both language and images. Although these combinations of language and images are a common feature of contemporary communication, relatively little is still known about their relationship with each other. For this reason, I aim to deconstruct the structure of the tourist brochures for a better understanding of how the brochures communicate using language, images, and other means of expression.

The kind of work undertaken in this dissertation is necessary, because written texts play a central role in everyday life. Consider, for instance, the average person whose daily routine begins with reading the newspaper in the morning. Yet a closer look reveals that the newspaper is far more than just a written text. Superficially, the written text in a newspaper is shaped by typography to organise the content and to help the reading process. Below the surface of the text is the linguistic structure, which varies according to what language is used for. Moreover, the newspaper is not exclusively verbal: the texts may be accompanied by photographs, graphics, diagrams, graphs and many other types of visual content. Finally, all of the above are brought together and organised in a layout.

It is precisely this interaction between the verbal and the visual that has caught the attention of linguists during the last two decades: how do the readers negotiate the combinations of verbal and visual content? What is the relationship between language and image? The search for answers has led to the emergence of multimodal research as a field of study. In short, multimodal research investigates the multiple modes of communication — language, image, layout, typography and many more — using methods informed by the field of linguistics. From a multimodal perspective, the written texts are much more than linguistic texts. All texts are inherently *multimodal* and for this reason, the term *artefact* may be more appropriate to describe these texts. This reflects a basic principle of multimodal



analysis: language is not alone at the centre of attention. Instead, all modes of communication are set on an equal footing.

Due to the growing interest within the field of linguistics, Forceville has suggested that the study of multimodality is currently a “hot academic topic” (2007, p. 1235). This is an important observation, because it is obvious that the fields of semiotics, media and communication studies, information design, art history and many others have previously studied the relationship and interaction between language and other modes of communication. Hence the question: how can the application of linguistic theories and methods enhance our understanding of multimodality, which in many aspects is a non-linguistic phenomenon?

## 1.1 Why use linguistic methods for multimodal research?

The field of linguistics has a long tradition of systematically studying the structure and functions of natural language, which is arguably the most complex mode of communication currently in use. This tradition presents a considerable advantage, because a growing body of research supports the argument that the multimodal artefacts do not simply use the visual elements to illustrate linguistic texts. Just as the structure of language varies according to what language is used for, so do the visual modes of communication and the structure of the entire multimodal artefact (see e.g. Martinec 2003; Machin 2004; O’Halloran 2005). Additionally, the sheer amount of multimodal artefacts in both print and digital media sets certain requirements for their study. To cope with the amount of data and the variation within them, we need applicable theories that provide consistent analytical methods to describe what takes place in the multimodal artefacts.

Bateman and Schmidt (2012, p. 32) have argued that the development of a “linguistically-inspired semiotics” for multimodal analysis offers many possibilities. They point out that the field of linguistics already possesses powerful analytic tools, which have been used to explore the many dimensions of language. In particular, linguistic research on the structure of discourse and its contextual interpretation may prove relevant for multimodal research. However, Bateman and Schmidt conclude that if these analytic tools are to be used in multimodal analysis, they need to be moved to “an appropriate level of theoretical abstraction” (2012, p. 32). This obviously anticipates the next question: which linguistic theories are applicable to multimodal analysis?

It is generally agreed that two intertwined theories of language have influenced multimodal research to a great extent: (1) social semiotics and (2) systemic-functional linguistics (see e.g. Kaltenbacher 2004; Martinec 2005; Jewitt 2009*a*).

This influence is evident in the early works of Kress and van Leeuwen (1990, 1996) and O’Toole (1994), which drew heavily on the aforementioned linguistic theories and are now considered seminal works in the field of multimodal research. Following the publication of these works, the subsequent research has branched off into three broad streams of research (Jewitt 2009a; Norris 2012):

1. *Social semiotic multimodal analysis* (Kress and van Leeuwen 2001, 2006)
2. *Multimodal discourse analysis* (O’Halloran 2004b, 2005, 2008b)
3. *Multimodal interactional analysis* (Norris 2004; Norris and Jones 2005)

At this point, however, it is not possible to enter into a discussion of the difference between the approaches or to showcase the wealth of research conducted within them. The relevant research will be covered in the literature review, which begins in Chapter 2. Right now, the discussion needs to focus on what can be done to move multimodal research forward.

To proceed, we may draw on the criticism presented towards multimodal research to identify the areas in need of development. So far, most of the criticism has targeted the social semiotic approach to multimodal analysis. For instance, Forceville (1999) pointed out at an early stage that the visual “grammar” proposed by Kress and van Leeuwen (1996) is less intersubjective than the authors suggest. Additional concerns have been later raised in Bateman et al. (2004) and Thomas (2009a), who have criticised the social semiotic approach as being “interpretative” and “impressionistic” particularly in the analysis of layout (see also Knox 2007, pp. 37-38). Recently, this criticism has also found support in information design, which is a field with considerable experience in the study of layout (Waller 2012, p. 243). Altogether, these concerns have lead to calls for increased empiricism in the field of multimodal research.

## 1.2 Why do empirical research on multimodality?

To bring out the benefits of doing empirical research, the current challenges to multimodal research need to be made clear. For instance, Forceville (2007, p. 1236) has indicated the problem of “infinite detail” in multimodal research: the performed analyses rarely formulate generalisations or make predictions about multimodality. Instead, they provide painstakingly detailed descriptions of carefully chosen artefacts or communicative situations. Given the immense number of multimodal artefacts in the world today, this presents a substantial challenge to the field, because a well-formulated theory should be able to generalise and make predictions about the studied phenomenon.

In terms of scale, Jewitt and Bezemer (2010, p. 194) have proposed that multimodal research is oriented towards “micro-interaction”, and thus the field cannot directly provide answers to questions about culture and society at large. Curiously, multimodal research has often sought to do exactly this by connecting the analyses to the broader issues of culture, history and society (see e.g. Machin and van Leeuwen 2005; Martínez Lirola and Chovanec 2012). Although many interesting topics may be raised in such analyses, they do little to advance our understanding of how the multimodal artefacts actually do their communicative work (Bateman and Schmidt 2012, p. 3).

Because the detailed analyses have dominated multimodal research, the more abstract levels of multimodal meaning-making have not been explored to a great extent, although they are likely to play an important role in organising the subject matter described by the detailed analyses (Lemke 2000). Moreover and regardless of the analytical level, multimodal analysis should adhere to the principles of solid theory-building, which Tseng and Bateman (2012, p. 93) succinctly outline for film:

[Analytic] schemes need moreover to *operate without specific commitments drawn from the individual film under analysis* in order to ensure comparability across analyses.

Their observation applies to the analysis of both static and dynamic multimodal artefacts. The capability to compare is essential, if multimodal research is expected to take on the wealth of artefacts currently in existence and to explain their specific characteristics and differences.

In order to meet this challenge, a step towards increased empiricism can be taken by sharpening the analytic tools used for multimodal analysis. This involves reworking the theoretical concepts, ideally by observing the data and feeding the results back into the theory (Bateman 2008, pp. 14-15). For multimodal research, the most important analytic tool is arguably the concept of *mode*, which describes how language, image, layout, etc. are used to make and exchange meanings (see e.g. Stöckl 2004; Kress 2009; Elleström 2010*a*; Bateman 2011). Another concept frequently deployed in multimodal research is *genre*, which is typically used to characterise multimodal artefacts or communicative situations, and their broader social and communicative purposes (see e.g. van Leeuwen 2005*b*; Baldry and Thibault 2005; Held 2005; Bateman 2008). Unfortunately, few attempts have been made to establish a connection between the fine-grained analyses of the modes and their relation to the abstract notion of genre.

The issues presented above may be brought together in two interrelated challenges to multimodal research: (1) the potentially excessive focus on detail and (2) our limited understanding of the abstract, higher-level organisation of multimodal meaning-making in artefacts. Together, these challenges pose a significant problem

for pursuing the general principles behind multimodality. As Tseng and Bateman (2012, p. 93) have proposed, the analytic schemes that are used to produce delicate multimodal analyses have to be anchored to the more abstract descriptions. Setting up a solid relation — based on empirical observation — between the different levels of abstraction identified in multimodal analysis is the only way to achieve a more comprehensive, general view of multimodality.

Such a relation should be established between the concepts of mode and genre, thus connecting the detailed and abstract descriptions. For this purpose, a good vantage point is provided by the multimodal artefact, which brings together the contributions from both mode and genre. This is precisely the area of multimodal research to which this dissertation aims to contribute by *modelling the structure of a multimodal artefact*.

The artefact may be thought of as the middle ground. From below, the notion of a mode provides a perspective to the multimodal structure and the particular selections made within the deployed modes. From above, the notion of a genre may be used to observe and identify patterns in the multimodal structure, which are typical to the artefact at hand. To be successful, however, the study of artefact structure has to be supported by a carefully-defined theoretical framework to bring the studied artefacts under analytical control. Moreover, an annotation scheme for creating multimodal corpora is required to handle the data.

By drawing on the methods of multimodal corpus linguistics, I attempt to provide a comprehensive view of the artefact's multimodal structure, beginning with the properties of the page as a physical object, asking what modes the page is licensed to deploy, and observing how the deployed modes are combined and configured on the page. I will now present the reasons behind undertaking such work in more detail.

### **1.3 Why model the structure of a multimodal artefact?**

Modelling the structure of a multimodal artefact benefits the entire field of study for several reasons, which are not limited to improving our understanding of the analysed artefact. Up to this date, multimodal analyses studying the artefact as a whole have been relatively rare (see e.g. Thomas 2009*b*). Yet such analyses hold considerable potential for theoretical development, as Forceville (2010, p. 2607) points out:

If we want to develop and refine ‘tools for analysis’, I suggest we do so by systematically analysing corpora of discourses (a) belonging to the same genre; (b) communicated in the same medium; (c) drawing on

the same combination of modes; (d) in the light of a clearly formulated research question.

Forceville's proposal may be understood as a call to complement the detailed analyses, which are strictly tied to their immediate analytical context, by 'zooming out' and also looking at multimodality on the more abstract levels. In doing so, we may learn in which contexts the available tools can still produce relevant analyses, and in which contexts these tools are found wanting.

Such a development, however, requires a carefully circumscribed data set and an appropriate set of analytical tools to support its analysis. If we take the notion of genre as the point of departure, it should be obvious that the artefacts belonging to a genre may take various forms (see e.g. Hiippala 2012*b*). The data set should therefore represent the variation that occurs in real life, instead of relying on hand-picked examples. Naturally, this variation sets a requirement for the analytic method: the chosen method has to be able to take on any instance of the data. And this interface between the data and the analytic method is right where the theoretical development and refinement may take place.

In this way, a model of artefact structure also makes a valuable theoretical contribution, as the model establishes a connection between the concepts of mode and genre. The model can help to situate the studied multimodal phenomena — such as metaphor, cohesion and intersemiotic relations (see e.g. Forceville 1996; Royce 1998; O'Halloran 1999*a*) — by painting the bigger picture of multimodal phenomena, while the close analyses contribute the fine detail.

To meet the criteria set out above by Forceville (2010), I decided to focus on a particular multimodal artefact with a specific topic: the tourist brochures promoting the city of Helsinki — the capital of Finland — published between 1967 and 2008. The motivation behind my choice will be presented in the following section.

## 1.4 Why study the tourist brochures?

The tourist brochures may be considered to have a dual function (Hiippala 2007, p. 10). On the one hand, the brochures provide the tourists with information about the destination and its surroundings, touristic activities, culture and so on (Valdeón 2009, p. 23). On the other hand, the tourist brochures are a form of marketing and advertising: they encourage the tourist to consume and perform the various activities associated with tourism (see e.g. MacCannell 1989; Jaworski and Thurlow 2009; Berger 2011).

Regardless of their informative or persuasive function, the tourist brochures are known to influence the image of a destination in the reader's mind, particularly through the visual modes of communication (Molina and Esteban 2006).

This observation, in combination with our knowledge of the linguistic structure of tourism discourse, supports the proposal that the tourist brochure is a multimodal artefact that is designed to do particular kinds of communicative work. These kinds of artefacts and their structure are the primary target of this dissertation, thus making the tourist brochures a suitable object of study.

### **1.4.1 The tourist brochure as a multimodal artefact**

What may be broadly defined as ‘tourism discourse’ has been studied in both linguistic and multimodal research. In this context, tourism discourse stands for the particular ways of using language and image to communicate information related to tourism (Thurlow and Jaworski 2010).

Previous research has established that tourism discourse has spread over print (Thurlow and Jaworski 2003; Hiippala 2012*a*) and digital media (Hallett and Kaplan-Weinger 2010). The discourse is often evaluative (Kaltenbacher 2007) and portrays the destination in a positive light (Hiippala 2007). Moreover, the artefacts participating in tourism discourse are inherently multimodal.

In this sense, the tourist brochures are a prime example of a multimodal artefact participating in tourism discourse. This may also explain why the brochures have been frequently studied in multimodal research. The multimodal structure of the tourist brochures has been previously explored in Yui Ling Ip (2008), Valdeón (2009), Francesconi (2011) and also in Hiippala (2012*b*), which was a pilot study conducted as a part of this dissertation. Additionally, potentially relevant work has been conducted within the fields of semiotics (Culler 1988; Edelheim 2007), cultural geography (Jokela 2011) and tourism studies (Scarles 2004; Molina and Esteban 2006; Garrod 2009). Looking at the previous research, it may be said that multimodality is an overarching theme in the analysis of the tourist brochures.

Despite the wealth of research, the previous work has not provided an exhaustive description of the tourist brochures as a multimodal artefact by asking why the brochures are structured the way they are? This is exactly the question that this dissertation attempts to answer to by developing a model of artefact structure. The research questions, which direct the dissertation in this task, are presented shortly in Section 1.5. Before presenting the research questions, I will briefly consider some commercial factors related to the study of the tourist brochures in Section 1.4.2 and their availability for multimodal research in Section 1.4.3.

### **1.4.2 Commercial factors and applications**

The work presented in Molina and Esteban (2006, p. 1051) shows that the tourism industry needs to understand how the tourist brochures work. They write:

It is thus important to establish some criteria for brochure design in order to adapt brochures to the specific needs of tourists and, consequently, to improve their appeal and efficacy in forming images ...

Molina and Esteban continue by pointing out that because the tourist brochures are heterogeneous in form, “it is necessary to define how brochures should be designed and what features they should have in common” (2006, p. 1051).

These features are multimodal, and for this reason, the brochure design is an area to which multimodal research can certainly contribute. In addition to the research presented above in Section 1.4.1, a model of the tourist brochure’s multimodal structure can bring new insights precisely in line with the requirements set out by Molina and Esteban (2006). By drawing on a multimodal corpus of heterogeneous data, it becomes possible to identify patterns in the tourist brochures’ multimodal structure.

There is also a considerable economic incentive to undertake this kind of multimodal research, because the worldwide tourism industry was valued at 740 billion euro in 2011 (World Tourism Organization 2012). Besides the possible applications, another reason to choose the tourist brochures as the data was their availability, which I will discuss next.

### **1.4.3 Access to research material**

The access to the research material also spoke in favour of choosing the tourist brochures as the data of this dissertation. In the case of the Helsinki tourist brochures, two organisations were the main sources of data: (1) the Helsinki City Archives and (2) the National Library of Finland. Both organisations store and maintain information on all types of documents and publications produced by the city of Helsinki. From these sources, I collected a data set that covers the period between the years 1967 and 2008.

There was also a concrete reason for acquiring the data from two different sources. For an unexplained reason, the Helsinki City Archives had lost the tourist brochures produced during the 1990s. Fortunately, these brochures were stored in the collections of the National Library of Finland. In addition, the most recent data were acquired by myself from the Helsinki City Tourist and Convention Bureau for previous studies (Hiippala 2007, 2012*a*).

A detailed description of the data collected for this dissertation will be provided in Chapter 5. I will now continue with the research questions of this dissertation.

## 1.5 Research questions and the structure of the dissertation

By drawing on the collected data, the dissertation attempts to answer several research questions, which are presented below in Section 1.5.1. After the presentation of the research questions, Section 1.5.2 outlines the structure of the dissertation.

### 1.5.1 Research questions

In this dissertation, I seek to answer the following research questions:

1. Which factors affect the structure of a multimodal artefact?

To understand how multimodal artefacts are construed, the various contributions to the artefact structure need to be clearly identified. Only then we may begin to consider how these contributions interact with each other.

2. What kinds of patterns may be identified in the multimodal structure?

It is reasonable to expect that multimodal artefact exhibit certain structural patterns, which are not limited to the linguistic structure. To establish how the differences between multimodal artefacts emerge, new analytic tools that go beyond the surface of the content need to be developed.

3. How does the structure of the multimodal artefacts change over time?

Change is an inherent feature of all modes of communication. For this reason, a model of a multimodal artefact needs to be able to pinpoint and explain which structures have changed over time and how.

These questions are considered imperative for modelling the structure of a multimodal artefact. By answering these research questions, we will stand in a stronger position to understand how multimodal artefacts work, both above and below the level of a page. We can then better situate the phenomena studied within multimodal research, which also opens up new possibilities of applying our knowledge in practice.

### 1.5.2 Dissertation structure

The structure of the dissertation is presented below. The theoretical framework is developed in Chapters 2, 3 and 4, followed by a description of the data and



methods in Chapter 5. Chapters 6, 7 and 8 present the analyses performed using the framework. Finally, the dissertation concludes with Chapters 9 and 10. I will now describe each chapter individually.

Chapter 2 discusses previous research on multimodality, explores parallel work in other fields of study, and establishes the major challenges faced by the field of study. The literature review is then followed by a discussion of the core theories of multimodal research and their applicability to the task at hand.

Chapter 3 describes the *Genre and Multimodality* model, which provides a model of artefact structure and an XML-based annotation schema for creating multimodal corpora. The notion of a semiotic mode is also considered, in order to provide the means to describe what exactly takes place in a multimodal artefact.

Chapter 4 considers the concept of genre, and how this concept may help to understand the content and structure of the multimodal artefact. In addition to the research on genre within the field of linguistics, the chapter introduces insights and recent advances from the fields of document theory and information design.

Chapter 5 presents the data collected for this dissertation, and how the data were annotated using the GeM model annotation scheme. The chapter also discusses the process of creating and verifying multimodal corpora, and presents the visualisations used in the later chapters.

Chapter 6 explores the brochure as a form of print media. The analysis, based on the collected data, addresses a variety of issues, such as the typical properties of the brochures, the role of advertising in the tourist brochures, and the much-discussed notion of a ‘visual turn’ in communication.

Chapter 7 demonstrates how the framework developed in this dissertation may be used to take apart the multimodal structure of the tourist brochures. Drawing on the multimodal corpus, the detailed analyses are used to identify specific patterns in the multimodal structure.

Chapter 8 takes a step back in the analysis and observes how the structural patterns identified in Chapter 7 are combined on the pages of the tourist brochures. These patterns are then used to establish the logic behind the organisation of a multimodal page and how this logic has changed over time.

Chapter 9 and 10 conclude the dissertation with a discussion of the results, convert these results into pragmatic advice for the tourism industry, evaluate the current study and propose several avenues of further research.

# Chapter 2

## Multimodality as a field of study

The second chapter of this dissertation has two goals: firstly, to establish the major theories of multimodality and to review their theoretical foundation, and secondly, to evaluate the capability of these theories to deconstruct, compare and explain the multimodal structure of the tourist brochures. The following sections work systemically toward these goals in the order described below.

Section 2.1 provides a point of departure for the theoretical framework by outlining the origins and development of multimodal research. Section 2.2 continues with a discussion of the influence of systemic-functional linguistics and social semiotics, followed by the key theoretical concepts in Section 2.3. Section 2.4 describes the multimodal analysis of print media, while Section 2.5 focuses on layout. The chapter concludes with final remarks in Section 2.6.

### 2.1 Origins and development

This section traces the development of multimodal research, follows some of the parallel work in other disciplines, and outlines the current challenges faced by the field, which together serve as the point of departure for developing the theoretical framework. I begin with the origins of multimodal research in Section 2.1.1. Section 2.1.2 continues with a discussion of related fields of study, such as semiotics, document design and visual rhetoric, and their relation to multimodal research. Finally, I conclude with a description of the current challenges faced by the field in Section 2.1.3.

#### 2.1.1 Origins

Kaltenbacher (2004) provides one of the first broad overviews of multimodal research, arguing that the meaning potential of semiotic resources and media has

been the subject of scholarly attention for a long time already. He cites the example of Gotthold Ephraim Lessing, a German classicist from the 18<sup>th</sup> century, who discussed the portrayal of Laocoön — a figure in Greek and Roman mythology — in an epic poem and in a sculpture by contrasting the properties of the semiotic resources used and their expressional capabilities and limitations. In hindsight, many aspects of research in art history, literature, semiotics, media and communication studies can indeed be considered multimodal. This observation, however, has only become possible after what Kaltenbacher refers to as “the revival or re-discovery of an important and interesting field of research” (2004, p. 192) that followed the publication of several influential works by Kress and van Leeuwen (1990, 1996) and O’Toole (1994).

These works drew on several theories of language — systemic-functional linguistics and social semiotics — to describe other forms of semiosis as communicative resources: image, composition, painting, architecture and sculpture, to name a few. Using the notion of resource to describe semiosis, as opposed to adopting rigid, rule-based descriptions, expanded these linguistic theories to an area of communication hitherto unexplored using linguistically-informed methods. Consequently, the works of Kress and van Leeuwen and O’Toole have contributed significantly to the theoretical discussion on multimodality and shaped the emerging field (Martinec 2005; Jewitt 2009*a*). However, it is also important to acknowledge and take heed of the previous contributions from other disciplines that have studied the foci of contemporary multimodal research, as they may offer new insights that have not received attention in the particular stream of multimodal research influenced by systemic-functional linguistics and social semiotics (see Section 2.2). In the following section, I present some examples of previous research from the fields of semiotics, document design and visual rhetoric.

### 2.1.2 Influences and parallel research

The pioneering work of Barthes (1977) on the semiotics of image-text relations has influenced multimodal research, because the co-deployment, interaction and interpretation of language and image constitute a major domain of research within the field. Indeed, as Forceville (2011*a*, p. 3624) points out, the work of Kress and van Leeuwen (1996) is indebted to Barthes, who was the first to argue that the analysis of images should not be limited to *artistic* images. Kress and van Leeuwen (1996, p. 17), however, refute Barthes’ argument that the interpretation of images always relies on language, and argue that both language and image undertake fundamental communicative tasks in society *independently* of each other, while remaining in simultaneous interaction. This leads us to the core idea of multimodality: although the contribution of language and image may differ across

communicative contexts, they should be set on an equal footing in analysis if we wish to understand how they work together.

The connections between the fields of semiotics and multimodality have been explored to some extent in subsequent work. For instance, Martinec and Salway (2005) integrate Barthes' image-text relations into their system developed for the same purpose, combining Barthes' work with the logico-semantic relations of Halliday (1985, 1994). In contrast, Kong (2006, pp. 208-209) argues that Barthes' concepts are outdated in the context of contemporary communication, reasoning that the complex relationship between language and images is not likely to be adequately explained by existing theories of language or image alone. As an alternative, Kong (2006, p. 211) proposes a taxonomy of logico-semantic relations for language-image relations, which draws on the corresponding linguistic model presented in Halliday (1994). It seems that the contribution of semiotics to multimodal research remains an open issue, as Aiello (2006, p. 100) suggests that multimodal research could still benefit from Barthes' work, particularly in describing the denotative qualities of image.

Nevertheless, Kong's statement may be true — at least in terms of detailed analysis and multimodal *structure* — considering Barthes' view that “analogical reproductions of reality”, such as photography and cinema, are “messages without a code” (1977, p. 17). In contrast, multimodal research has provided growing evidence that both photography (see e.g. Machin 2004; Caple 2009a; Caple and Bednarek 2010) and cinema (see e.g. Bateman 2007; Tseng and Bateman 2010; Bateman and Schmidt 2012; Wildfeuer 2012) are indeed highly structured and ‘codified’. Moreover, the notion of a ‘code’ has been contested by van Leeuwen (1999, pp. 4-5), who advocates using the notion of ‘resource’, because ‘code’ implies a static and non-dynamic entity. This argument is supported by recent research, which suggests that semiotic resources are inherently dynamic (see e.g. O'Halloran 2009a; Zhao 2010a). To sum up, Kong's observation has two implications for multimodal research. Firstly, there is a need to actively develop new theoretical frameworks, and secondly, these new theories need also to be subjected to a critical examination and tested empirically. In this aspect, this dissertation directly contributes to these causes by evaluating the GeM model (Bateman 2008) and its applicability to modelling the structure of a multimodal artefact.

However, in order to avoid ‘reinventing the wheel’, it is also necessary to consider the contribution of some other fields with shared research interests in multimodality, although their metalanguage and vocabulary may differ from that of contemporary multimodal research. For instance, Twyman (1979, 1986) developed methods for describing the spatial configuration of verbal and visual elements in print media and traced their development over time. Waller (1987), in turn, explored the contribution of typography in multimodal artefacts (for work influenced

by Waller, see Bateman 2008; Thomas 2009*b*), while Eiler (1987) offered one of the first social semiotic perspectives to document design, describing how “design maxims” arise from a text and its communicative function. In addition, contemporary research in visual rhetoric by Kostelnick and Hassett (2003) has explored the role of design conventions in multimodal artefacts. Finally, in the field of experimental psychology, Hegarty and Just (1993) and Hannus and Hyönä (1999) have explored the role of multimodality in learning and the transfer of knowledge.

To sum up, even a brief glance at the previous research in different disciplines underlines the complex nature of multimodality as a phenomenon whose understanding is likely to benefit from an interdisciplinary approach. For this reason, this dissertation pays special attention to the relevant research in other disciplines, while maintaining that interdisciplinary work should be a two-way street: it is also important to consider what multimodal research can contribute.

Some examples of interdisciplinary work in multimodality include Bell (2001) and Bell and Milic (2002), who combined the content analysis of Goffman (1979) with a social semiotic framework. More recently, Holsanova and Holmqvist (2006), Holsanova and Nord (2010), Boeriis and Holsanova (2012) and Hiippala (2012*c*) have explored the research of visual perception in experimental psychology and how this research may inform our understanding of the perception of multimodal artefacts. It appears that interdisciplinarity is emerging as a key factor for meeting the current challenges in multimodal research by providing novel perspectives from outside of the field (cf. Luke 2003; van Leeuwen 2005*c*; Kress 2011). These challenges are discussed in the following section.

### 2.1.3 Current challenges

It may be suggested that some of the challenges facing multimodal research do not only result from the inherent multimodality of human communication and our need to understand it, but also from the broad developments currently taking place in the society, which shape the way we communicate with each other. As O’Halloran et al. (2010, p. 3) write:

The twentieth century was a time of rapid change and growth in the study and understanding of human meaning systems. Whereas scholars prior to the mid twentieth century were on the whole in the relatively safe position of being able to restrict themselves to the study of (for the most part written) language, the ongoing revolution in multimedia design and digital technology within contemporary society at large has led to a proliferation of multimodal documents (using media such as graphics, digitized photographs, audio and video texts, contemporary digital media, in particular web-based media, and three-dimensional virtual

objects). Meanwhile, the twentieth century has seen a profusion of disciplinary and theoretical perspectives upon human communication in general and multimodal communication in particular.

The challenge is thus presented by new and increasingly complex forms of semiosis across a range of different media. As I pointed out above, multimodal research explores semiosis using linguistically-informed methods, which underlines the rapid development of disciplinary and theoretical perspectives since Barthes proclaimed that “linguistics stops at the sentence” (1977, p. 82).

According to Kong (2006, p. 209), the proliferation of communicative situations and artefacts involving multimodality has led to the point whereby existing theories have to be refined for a more accurate description of the studied phenomena. However, the challenge presented to multimodal analysis by information technology may also provide solutions to analytical problems within the field. In recent years, significant advances have been made in incorporating information technology in multimodal analysis: O’Halloran (2008*b*) has used graphics editing software to analyse artefacts in static media, while Tan (2009) has demonstrated how software can be used to identify and deconstruct complex multimodal phenomena involving moving images. O’Halloran et al. (2010, 2011, 2012) and Smith et al. (2011), in turn, have continued the discussion of designing software for multimodal analysis and its implications to the field.<sup>1</sup> Finally, practical applications of information technology to multimodal analysis have been exemplified by the automatic visualisations described in Thomas (2009*b*) and Podlasov et al. (2012).

In linguistics, information technology has contributed significantly to the rise of corpus linguistics and provided the means for increased empiricism (see e.g. Biber 1988; Leech 1991). Teich (2009, p. 113) sums up the development succinctly: “linguistic work is unthinkable today without the support of computers”. This observation applies increasingly to research in multimodality as well, as corpus-driven methods have been carried over to multimodal research, for example, in the work of Baldry and Thibault (2005), Baldry (2007), Bateman (2008), Velasco (2012), Taboada and Habel (2013) and Kong (2013). Additionally, the potential applications of multimodal corpora have also been examined by Thomas (2007, 2009*b*) and Hiippala (2012*c*), who have explored the possibilities of interfacing Bateman’s (2008) model with concordancers, optical character recognition and eye-tracking equipment. However, Parodi (2010, p. 72) notes that although standard markup languages such as XML may provide the backbone for multimodal corpora, work remains to be done in developing theories of multimodality that would enable the researchers to harness the full potential of corpora, especially in the area of automatic segmentation and annotation of data (see e.g. Allwood 2008; Flewitt et al.

---

<sup>1</sup>This work has later resulted in commercial applications for multimodal analysis, see [multimodal-analysis.com](http://multimodal-analysis.com).

2009). Future work is therefore likely to involve developing efficient means of creating corpora, and improving theory-building by using corpora to test hypotheses and feeding these findings back into the theory.

One domain of multimodal research where a close relationship between theory, data and information technology is currently emerging is the analysis of film and television. In this domain, several research streams have appeared following the early multimodal investigations (see e.g. van Leeuwen 1991, 1996; Thibault 2000). To begin with, the development and application of computer tools for multimodal analysis has been particularly strong in the aforementioned work of O'Halloran (2004*a*), Tan (2009) and O'Halloran et al. (2010, 2011, 2012). Another recent development has involved multimodal descriptions of the various aspects of 'filmic' semiosis, which also draw on film studies and document theory (Bateman 2007, 2009*c*; Tseng 2008, 2009; Tseng and Bateman 2010, 2012; Bateman and Schmidt 2012).

The ongoing work on the analysis of film and television has also contributed to the analysis of digital media, which allows dynamic content in the form of moving images. Digital media has also received increased attention in multimodal research, most likely due to its growing role in contemporary communication. Some examples of recent research include websites (Kok 2004; Djonov 2007; Bateman et al. 2007; Knox 2007; Hopearuoho and Ventola 2009; Tan 2010), presentation software (Djonov and van Leeuwen 2011*a,b*), digital art (Huemer 2008) social networking sites (Eisenlauer 2011) and online games (Maiorani 2009; Stenglin and Djonov 2010).

In this connection, it should be noted that the major theories of multimodality have emerged during a period of rapid technological change in the last 20 years. During this period, the role of information and communication technology has grown in terms of both availability and everyday use. This development has provided the multimodal analysts with a wealth of new situations and phenomena to study, but concerns about the theoretical consequences of the development have also been raised. As Bateman (2008, pp. 213-214) writes:

... a word of caution for the currently exploding 'study of websites' being undertaken in multimodal linguistics: there are useful and interesting aspects of websites to investigate, but they are found exceedingly rarely directly on the surface in terms of novel design and new genres. To what extent the web-page has moved beyond technological dependence to support semiotically interesting meaning-making possibilities is an open question, a question that is not addressed adequately simply by assuming it to be the case.

Increased attention should therefore be paid to the development of empirically-oriented frameworks for the analysis of digital media. However, before advancing

to the description of dynamic digital media, we need to have a solid understanding of the basic principles of multimodal meaning-making, to which this dissertation contributes by modelling the structure of a static artefact. An understanding of the basics is also a prerequisite for applying the newly acquired knowledge of multimodality, as Livingstone (2004, p. 12) points out:

Until we have a robust account of the media in which people might be judged literate, we can say little about the nature or uses of their literacy.

In short, an assumption that a rapid change in semiosis has taken place should not be made on the basis of transition from one medium to another. Language and other semiotic resources are unlikely to change at such a rapid pace, because their users would be unable to keep up with the change. As I will show in this dissertation, tracking the changes in the functions and structure of the semiotic modes is necessary for describing multimodal communication in the world today. By observing the past, we may better understand how the present situation emerged.

To conclude, the state-of-the-art in multimodal research may be summed up in two points. Firstly, the increasing use of information technology should be accompanied by a simultaneous drive to develop and redefine the theoretical frameworks to a degree where the performed analyses (1) may be replicated using different data and (2) are not limited to their immediate analytical context. Secondly, research evidence from outside the field of multimodal research should not be ignored simply because it does not have the same underlying theoretical principles and a compatible metalanguage (see Section 2.1.2). By combining information technology and corpora with a carefully defined theoretical concepts, this dissertation accommodates both of these points and brings them together in a model of artefact structure.

Finally, despite the increased availability of analytical and technological tools, the theories that comprise the theoretical foundation of contemporary multimodal research may have received less attention than they warrant (cf. Zhao 2010*a*). I argue that this is especially reflected in the systemic-functional and social semiotic theories and their extension to multimodal research, in which the same theoretical concepts are often applied to describe different semiotic resources. In any case, the contribution of social semiotics to the study of multimodality has been and continues to be remarkable, and therefore an understanding of social semiotics and the associated linguistic theories is a prerequisite for a theoretical discussion of multimodality. For this reason, I shall now move to discuss the influence of systemic-functional linguistics and social semiotics on the theories of multimodality.



## 2.2 Influence of systemic-functional linguistics and social semiotics

The fields of systemic-functional linguistics (hereafter SFL) and social semiotics have significantly influenced the development of multimodal theories (Jewitt 2009*a*, p. 36). This is widely acknowledged in the overviews of multimodal research as a field of study (Kaltenbacher 2004; Martinec 2005; Jewitt and Bezemer 2010; O’Halloran 2011). Therefore, the purpose of this section is to provide a sufficient introduction to both approaches, so that their origins and contribution to multimodal research may be acknowledged and subsequently evaluated.

To begin with, SFL has its roots in Firthian linguistics (see e.g. Firth 1957), developed by the linguist Michael A. K. Halliday into two theories:

1. *Systemic-functional linguistics*: a theory of linguistic structure (Halliday 1976*b*, 1985, 1994; Halliday and Matthiessen 2004).
2. *Social semiotics*: a theory of language in society (Halliday 1978).

This research has expanded into multiple research streams, whose individual descriptions lie beyond the scope of this dissertation (for social semiotics, see e.g. Hodge and Kress 1988; van Leeuwen 2005*a*; Kress 2010). Comprehensive overviews of SFL research have been presented in Monaghan (1979), Butler (1985), Fawcett (2000) and Martin (2011), while accessible introductions to the SFL theory have been provided by Butt et al. (2000), Thompson (2004) and Eggins (2004). Additionally, the central theoretical concepts have been explained in Halliday (2009).

As a review of the multimodal research will show, the broad social perspectives of SFL and social semiotics on language and its functions have provided these theories with the potential of being applied to the description of other semiotic resources. Consequently, this dissertation cannot ignore the influence of SFL and social semiotic research, their models and theoretical concepts. Moreover, the data of this dissertation which combines language and image on printed pages represents the kind of complex multimodal phenomena that would not likely be adequately explained by linguistic models with a narrow perspective. At the same time, however, it is also necessary to avoid “linguistic imperialism” and the overextension of linguistic theories and concepts (cf. Scollon and Scollon 2009). However, in relation to multimodal analysis of film, Bateman and Schmidt (2012, p. 32) suggest that:

Linguistically-inspired semiotics ... has much to offer precisely because linguistics as a science has now explored many of the semiotic dimensions necessary in considerable detail.

The wealth of multimodal research in recent years shows that linguistics can contribute to the understanding of semiotic phenomena *beyond* language. The

important question is, which concepts and theories may be applied and how, and how far can they take us in terms of producing new and reliable descriptions of multimodal phenomena? Machin argues that SFL and social semiotics “hold the promise of facilitating a more systematic way to analyse visual communication which has been largely dominated by more general open interpretation” (2009, p. 183). However, concerns have been also raised about the frequent borrowing of theoretical concepts and its consequences to the development of multimodal theories (see Zhao 2010*b*). To tackle the issue of theoretical concepts and their application, the following sections shall look at the underlying theoretical assumptions in SFL and social semiotics, in order to establish whether these theories can provide the tools for systematically investigating the structure of a multimodal artefact in the way envisaged by this dissertation.

### 2.2.1 Overview of the underlying theories

To begin with the basic principles of SFL and social semiotics, Halliday (1987/2003, p. 117) sees language as an evolved system, which is employed by its users in situations involving interaction and the exchange of meanings in all aspects of social life (cf. Halliday 1978). SFL and social semiotics therefore place special emphasis on the functions of language in social contexts, and make the use of language in context their prime target of investigation (cf. Martin 1991, 1999). Language, the semiotic system that provides the meaning potential harnessed in social situations, is modelled in SFL as an interconnected series of choices made simultaneously in multiple *systems*. These choices are represented visually using system networks: an example network is shown in Figure 2.1.<sup>2</sup> This lends the theory its epithet ‘systemic’, while the term ‘functional’ arises from emphasis on the functions of language.

With the basic concepts now introduced, the following quote from Halliday and Matthiessen (2004, p. 23) may help to establish an overall picture:

A text is the product of ongoing selection in a very large network of systems — a **system network**. Systemic theory gets its name from the fact that the grammar of a language is represented in the form of system networks, not as an inventory of structures. Of course, structure is an essential part of the description; but it is interpreted as the outward form taken by systemic choices, not as the defining characteristic of language. A language is a resource for making meaning, and meaning resides in systemic patterns of choice.

Furthermore, SFL considers the relationship between language and text as a cline, where the meaning potential of a language is realised in a particular instance of a

---

<sup>2</sup>For elaborate examples of the system networks in the English language, see Halliday (1976*a*).

text. Lukin et al. point out that this cline largely solves “Saussure’s unnecessary bifurcation of *langue* and *parole*” (2008, p. 193), which are simply two different points of observation. In a similar way, Halliday and Matthiessen (2004, p. 26) propose that the relationship between language as a system and language as a set of texts is analogous to the relationship between climate and weather: the two are essentially the same phenomenon, but observed from a different standpoint. The weather is an instance of the climate, while the text is an instance of the underlying system.

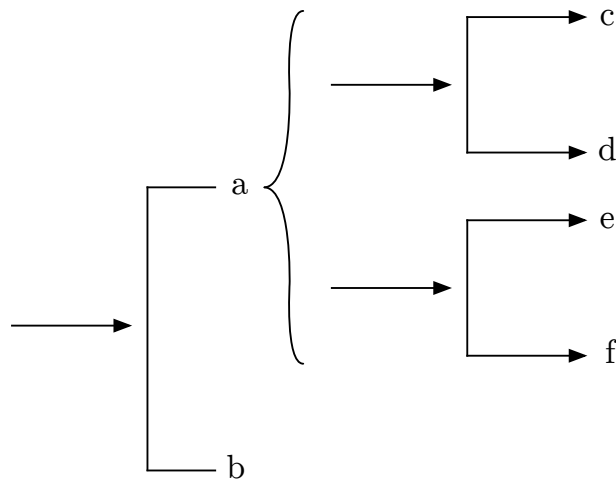


Figure 2.1: A system network with an entry condition and a binary choice between either *a* or *b*. The curly bracket implies simultaneous choice: if *a* is chosen initially, then a further choice must be made between both *c* and *d* and *e* and *f* (see also Eggins 2004, pp. 194-200).

The ‘end products’ of the system — the texts and their linguistic features — have been described in SFL using the notions of genre and register. I will discuss both genre and register later in Section 2.3.1.1, and thus a brief description that applies to both should suffice for the time being: these features are considered to emerge from the context and function of language use. To illustrate the different contexts of language use and the ensuing linguistic variation, Halliday and Matthiessen (2004, p. 27) list a small sample of innumerable situations where language is used: recipes, stock market reports, walking tours in a guide book and advertisements, to name a few. The list proves endless, especially when including the situations in which spoken language is used (see e.g. service encounters in Ventola 1987). In all of these situations, language acts as a social semiotic, which is used to make and exchange meanings: this provides the basis for interaction between humans (Halliday 1978, pp. 21-22).

In this aspect, SFL adopts a different approach to language compared to — for instance — the transformational grammar pioneered by Chomsky (1957) and the subsequent work, which decontextualises language in order to formulate a set of rules for producing syntactically correct sentences (Chomsky 1995, 2006). Halliday and Fawcett (1987, pp. 1-2) point out that the rules produced by transformational grammar often describe a limited area of syntax, leaving little room for cooperation and building a theory that would adequately describe the breadth of language as a phenomenon. SFL and social semiotics, in turn, aim to account for the vast phenomenon of language by suggesting that the context of use is a major factor in shaping language, proposing that language should not be seen as a code or a set of rules, but as a *resource* and an object of study, which may be described *using* language (Halliday 1981, p. 16).

This brings us to the final point. In the terms of Hasan (2005, pp. 51-52), SFL and social semiotics can be considered “exotropic” theories, that is, they seek a dialogic engagement with other fields on inquiry. However, Hasan also argues that the core logic of an exotropic theory needs to be well-established before its extension to other fields. As I pointed out above, Zhao (2010*b*) has raised concerns about the nature of the dialogue between SFL, social semiotics and multimodality, which needs to be accounted for in the theoretical framework developed for this dissertation. Otherwise, the time appears to be right for applying linguistic frameworks to the analysis of multimodality, as we now possess frameworks that are sufficiently developed in terms of their ‘core logic’ to describe other forms of meaning-making (for an extensive discussion, see Bateman and Schmidt 2012, Chapter 2). I will therefore continue to explore the dialogue between systemic-functional and social semiotic theories and multimodality below.

### 2.2.2 Implications to multimodality

Given the broad outlook of SFL and social semiotics on language, it is not surprising that the interest in multimodality gained currency especially within these fields. Whereas SFL is geared towards the study of language, theories of multimodality consider language as just one of the semiotic resources used in the exchange of meanings, thus expanding the study of meaning-making to other semiotic resources and the interaction of these resources. From a systemic-functional perspective, a multimodal text is an instance resulting from choices made in multiple systems (and systems within systems) in multiple semiotic resources (cf. O’Halloran 1999*a*, 2004*a*, 2008*b*; Liu and O’Halloran 2009). The list of text types in Halliday and Matthiessen (2004, p. 27) may also be used to highlight the complexity of multimodal meaning-making. In contemporary procedural texts, such as recipes, images are used together with language to portray sequential action (Martinec 2003), while

advertisements use both language and images to persuade their audience (Cheong 2004; O'Halloran and Lim 2009; White 2010; Berazhny 2012).

In these cases, multimodal complexity does not only result from the mere presence of both verbal and visual elements and their visual-spatial organisation as a part of a layout, but from the delicate choices in typography (font face, size, weight, etc., see e.g. Stöckl 2005; van Leeuwen 2005*d*, 2006), colour (tone, contrast, etc., see e.g. Kress and van Leeuwen 2002; van Leeuwen 2011) and image (camera angle, distance, lighting, etc., see e.g. Machin 2004; Lim 2004*a*), and particularly from the interaction of the semiotic resources: cohesive ties, metaphor, image-text relations and so on, many of which have been described using frameworks that draw on SFL and social semiotics (see e.g. Kress and van Leeuwen 1996, 2006; Royce 1998, 2007; O'Halloran 1999*a*; van Leeuwen 2005*a*; Martinec and Salway 2005; Kong 2006; Liu and O'Halloran 2009; Kress 2010; Chen 2010*a*; Guijarro 2011).

To conclude, the aforementioned multimodal phenomena may also be expected to be encountered in many different forms in the data of this dissertation. Most importantly, to capture the multimodal structure of the tourist brochures, the dissertation needs a comprehensive and analytically reliable set of theoretical concepts. Only by bridging the gap between these concepts and the notion of structure, it becomes possible understand how complex multimodal phenomena emerge. This is also why the close relationship between SFL, social semiotics and multimodal research requires a closer and critical examination, which focuses especially on the influence of SFL and social semiotics on the development of multimodal theories. In short, it is necessary to evaluate whether these theories can provide the tools for accomplishing the goals of this dissertation. The following section thus presents the key theoretical concepts that have been carried over to multimodal research and discusses how they may inform a model of artefact structure.

## 2.3 Key theoretical concepts

In this section, I discuss the core theoretical concepts of multimodal analysis that have their origins in the systemic-functional and social semiotic theories of language. It should be noted, however, that these concepts are in a state of constant flux due to the ongoing theoretical development (Jewitt 2009*b*, p. 21): I therefore aim to provide a view of their foundation. The three theoretical concepts described in this section are *mode* in Section 2.3.1, *rank scale* in Section 2.3.2 and *metafunctions* in Section 2.3.3.

## 2.3.1 Mode

The notion of mode is critical to any investigation of multimodality, as it provides the means to discuss the contributions of language, image, colour and other semiotic resources to a communicative situation or artefact. So far, I have simply used the term semiotic *resource* to describe the different contributions to communication. It may be suggested, however, that the concept of a mode should be at the heart of any theory of multimodality. I thus dedicate the following sections to an in-depth description of the semiotic modes: I will also clarify the difference between a semiotic resource and a mode below.

What exactly constitutes a mode has been debated extensively in multimodal research, which is evident in the range of different definitions provided in recent work (see e.g. Kress 2009; Elleström 2010a; Bateman 2011). The following sections thus work towards the goal of describing the notion of mode in the following order. Section 2.3.1.1 traces the development of mode as a part of the related concept of register, before moving to discuss mode in multimodal research in Section 2.3.1.2. Section 2.3.1.3 describes the relationships between different modes, whereas Section 2.3.1.4 discusses the influential social semiotic concept of mode.

### 2.3.1.1 Mode in linguistics

I shall begin with a linguistic definition of mode. According to Halliday (1975, p. 131), mode is one of the three domains that constitute the “environmental determinants of text”: the other two are field and tenor.<sup>3</sup> Field refers to the subject matter, or what is being talked about, whereas tenor reflects the social relations between the speakers and their effect on language use. Together, the field, tenor and mode are known as the *register* variables: they form the backdrop for language use and collectively determine the kinds of selections made in language (Halliday 1978, p. 223).

For Halliday (1978, pp. 222-223), the concept of mode covers two aspects, which are illustrated in Figure 2.2:

1. The rhetorical mode of language in use.
2. The selected medium of communication.

The rhetorical mode structures language according to the task it is being used for. Halliday (1978, p. 222) underlines this point by showing how the structure of language varies in, for example, when it is used to sell or to teach. Additionally, the selection of the medium affects the structure as well: Halliday (1975, p. 133) uses

---

<sup>3</sup>Due to their role as functions, Field, Tenor and Mode are commonly written with a capital first letter in SFL literature, this convention is not followed in this dissertation due to its focus on multimodality.

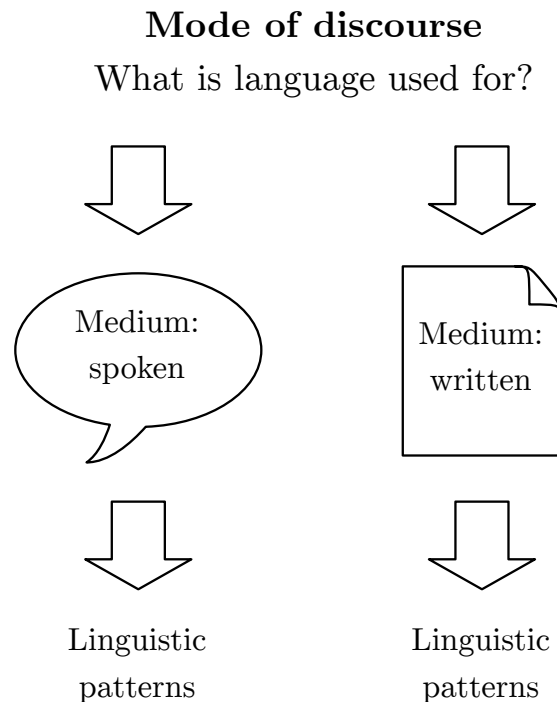


Figure 2.2: Linguistic mode in Halliday (1978)

spoken dialogue as an example of how the used medium determines the linguistic patterns typical to the situation in ellipsis, choices of voice and theme, lexical continuity and exophoric reference. The alternative mode for language is the written language, which features other types of linguistic patterns.

At this point, it is important to distinguish between the alternative definitions of register as a theoretical concept. Lukin et al. (2008) provide a comprehensive overview of the issue, describing the differences between Halliday’s approach to register and the use of register in the work of Martin (1992).<sup>4</sup> The major difference between the two accounts is that Martin’s model extends the scope of investigation far beyond the analysed text, positioning the additional strata of genre above the register variables (for a discussion of genre, see Section 4.3). According to Lukin et al. (2008, p. 208), the problem is that the difference between the concepts of register — and mode — has become obscured in subsequent work.

Lukin et al. (2008, p. 188) also quote Hasan on a matter which is particularly relevant to multimodal theories of mode and this dissertation: “a theoretical term ... requires that it be positioned *vis-à-vis* other concepts in the theory” (2004, p. 16). To put it simply, it is also important to consider how the theoretical concept

<sup>4</sup>For accounts of the work pursued in this research stream, see Martin (1999, 2009).

fits into the overall theory of multimodality and how the concept interacts with other theoretical concepts. In modelling the structure of a multimodal artefact, all the pieces matter, and thus the compatibility between the concepts that describe the different pieces warrants special attention.

### 2.3.1.2 Mode in multimodal research

From the perspective of SFL and social semiotics, the theories of multimodality propose that in many forms of communication, mode involves multiple interacting semiotic resources: this is why communication is considered *multimodal* (cf. Martin 2009). Consider, for example, the previous example from Halliday (1978, p. 222) on how the structure of language varies when it is used to sell or teach. From a multimodal perspective, similar variation may be observed in the configuration of other semiotic modes — such as image — in advertising (see e.g. Forceville 1996; Cheong 2004; Maiorani 2008; Hopearuoho and Ventola 2009; White 2010) and in education (see e.g. O’Halloran 1999*a*; Guo 2004; O’Halloran 2008*a*; Chen 2010*a,b*). Moreover, this variation takes place on various levels, and therefore it may be suggested to be stratified. Uncovering the factors that affect the configuration of semiotic modes on various strata and in different contexts is necessary for understanding multimodality: acknowledging their contribution and multimodality as a phenomenon is simply the starting point. With carefully defined theoretical concepts, we can already make considerable progress in solving many parts of the puzzle, as I aim to show in the subsequent chapters.

The multimodal patterns and their formation in artefacts and situations are also of high interest to this dissertation and therefore deserve special attention. For this task, the dissertation requires solid notions of a semiotic *resource* and a semiotic *mode*. As Jewitt (2009*b*, p. 21) has noted, the definition of mode is problematic, as the concepts of mode and semiotic resource are understood differently by the various approaches to multimodality. Following Bateman (2008, 2009*b*, 2011), this dissertation considers language and image as semiotic resources, which may participate in semiotic modes. How exactly these two concepts work together will be explained in Section 3.3. The remainder of this section, in turn, will evaluate the possible contribution of alternative definitions of mode. Note that the division between a semiotic resource and mode used in this dissertation does not apply in the following discussion. For the sake of clarity, the concepts are used as originally defined by their authors.

### 2.3.1.3 Modal interrelations

I begin with the work of Stöckl (2004), who has explored modal interrelations in print media to a great extent. This work is particularly relevant, because under-



standing the interaction of the semiotic modes in print media is a major concern for this dissertation. The model presented in Stöckl (2004) is given in Figure 2.3, which shows where the decisive choices that affect the properties of a semiotic mode are made, and how these choices relate to each other.

First of all, Stöckl (2004, p. 11) points out that mode cannot be defined according to the *sensory* channel (visual, auditory, tactile, olfactory or gustative) used to perceive a sign, because these categories are too broad for a theory of multimodality. Language, for example, would fall into both visual and auditory channels, depending on whether the language is written or spoken; similar views have been later presented in Elleström (2010a) and Bateman (2011). To overcome the problem of sensory channels, Stöckl proposes a model of print media in which language and image constitute the core modes. Typography and layout are described as peripheral modes of *writing* (Stöckl 2004, p. 12), which is the written medial variant of language. In this medial variant, the core mode of language has several peripheral modes and sub-modes, which act as its building blocks: font face, size, weight, colour, paragraphing, margins, etc. The typographic and graphic features of the sub-modes, in turn, shape these building blocks.

On a more abstract level, Stöckl suggests that “genres or text types differ a great deal in how they foreground and background modes and sub-modes” (2004, p. 15). This observation makes an important contribution to modelling the structure of a multimodal artefact in two ways. Firstly, Stöckl captures the elastic nature of the semiotic modes in a multimodal artefact: the configuration of the semiotic modes, “foregrounding and backgrounding”, bears close resemblance to what Bateman (forthcoming) calls a rhetorical strategy. Secondly, the relationship between a mode, its sub-modes and their features deserves attention, because it enables this kind of elastic configuration to take place. Together, these factors — the flexible deployment and configuration of modes — constitute an important aspect of describing a multimodal artefact. However, to pull these factors apart, the analysis needs to be supported by a sufficiently robust notion of multimodal structure. A candidate for this task will be presented in Chapter 3.

In addition, Stöckl (2004, pp. 16-18) suggests that the core modes may be distinguished according to three different characteristics. First, the semiotic properties of a mode have to do with its organisation: whereas the structure of language is redundant, image does not possess distinct units. The second level is that of cognitive orientation, which contrasts the linear organisation of language to the holistic nature of images.<sup>5</sup> Third, Stöckl identifies the level of semantic potential, that is, what the mode can be used for. He then makes a particularly interesting point regarding the difference between language and image (2004, p. 17-18):

---

<sup>5</sup>However, images can take part in a linear organisation: see Section 3.4.2.

Channels, modes and sub-modes in printed media

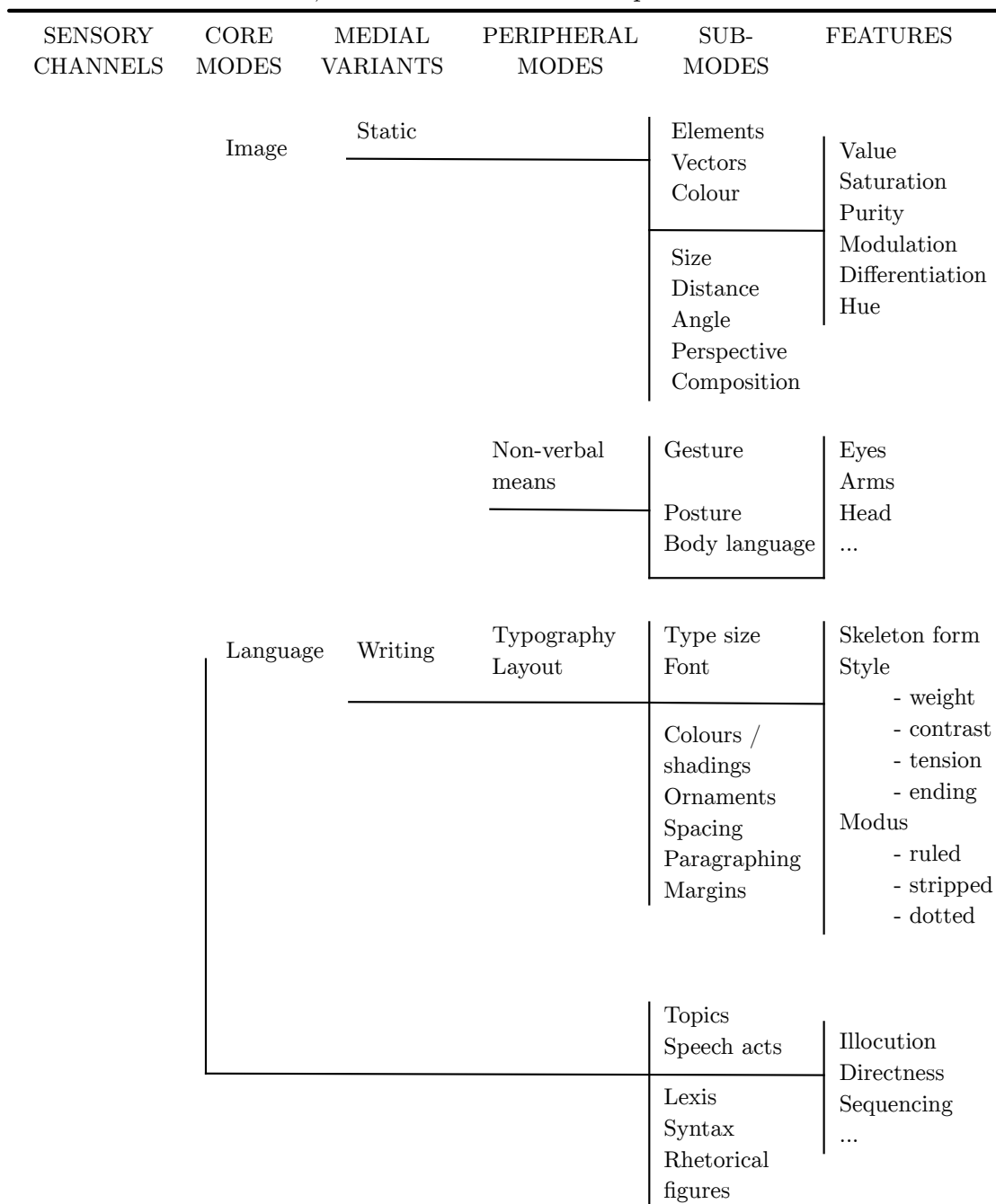


Figure 2.3: Network of modes, sub-modes and features in printed media  
(Reproduced from Stöckl 2004, p. 12)

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 the other modes or the embedding into narrowly defined communicative situations.

Here Stöckl touches upon an important aspect of the semiotic modes, that is, how we make sense of language and image in different contexts, and how these situated discourse interpretations emerge (see Section 3.5.3). Decomposing the structure of a multimodal artefact also requires an investigation of what the semiotic modes do in the particular artefact, in order to explore the relationship between the configuration of the semiotic modes and their communicative functions. As I will show below, this requires a consideration of how the core modes work together.

Overall, the work of Stöckl (2004) provides many valuable perspectives to the work undertaken in this dissertation by outlining the factors that affect the structure of a multimodal artefact. In some cases, however, the focus needs to be sharpened to make a contribution viable. A particular concern is the foregrounding and backgrounding of the semiotic modes in different genres or text types: before making such statements at such an abstract level, we should consider how these processes operate on the different *pages* of an artefact. Although I will now take a step back and observe a broader definition of mode in the social semiotic tradition, the issue of page types will be taken up again in Section 2.4.

#### **2.3.1.4 Mode in social semiotics**

Adopting a social semiotic perspective to mode, Kress and van Leeuwen suggest that “modes are semiotic resources which allow the simultaneous realisation of discourses and types of (inter)action” (2001, p. 21). By emphasising that modes are shaped in interaction with their surroundings, Kress and van Leeuwen (2001) relate mode to the concept of design within a larger group of processes governing multimodal communication: discourse, design, production and distribution. The designs formulate and combine discourses by tapping into the semiotic potential of the modes, which are exploited according to the intended function of the design (Kress and van Leeuwen 2001, pp. 20-21). The modes are thus considered to be socially and culturally shaped (Kress 2009, p. 54), which characterises the social semiotic approach to multimodality (Jewitt 2009a, pp. 30-31) and provides a rich theoretical framework for a detailed discussion of multimodal phenomena (see Section 2.4.2).

Recently, the broad definition of modes as semiotic resources in Kress and van Leeuwen’s framework has been criticised by Elleström (2010a, pp. 40-41), who

argues that mode has not been sufficiently theorised as a part of the work of Kress and van Leeuwen (1996, 2001, 2006) and in subsequent work such as that of Bateman (2008). As a result, the multimodal research following this tradition lacks an adequate definition of mode. According to Elleström (2010*b*, p. 14), the contribution of modes is difficult to analyse and compare due to their overlapping features. Elleström (2010*b*, p. 36) proposes that a more accurate description can be achieved by defining four separate *modalities*: the material, sensorial, spatiotemporal and semiotic modalities, all of which consist of several modes. For example, the sensorial modality includes the modes of seeing, hearing, feeling, tasting and smelling, whereas the spatiotemporal modality is concerned with the modalities' use of spatial dimensions. To exemplify, a printed photograph is two-dimensional, but a dance involves three spatial dimensions and the dimension of time (Elleström 2010*b*, p. 19).

Nevertheless, despite an important contribution to the theoretical discussion on mode — and especially on how they may be deconstructed — Elleström neglects the work of Lim (2004*a,b*) and Stöckl (2004), where the concept of mode has been extensively discussed, and the more recent work of Bateman (2009*b*, 2011). In several aspects, Elleström's work also overlaps with that of Bateman (2008, 2009*b*, 2011) and Bateman and Schmidt (2012), particularly in the discussion of “basic” and “technical” media (Elleström 2010*b*, p. 33), which mainly concerns the constraints that a medium sets on the deployment of semiotic modes. These constraints are indeed a significant issue, which I will cover extensively in Chapter 6.

Finally, let us consider the definitions of mode provided by Stöckl (2004), Kress and van Leeuwen (2001) and Elleström (2010*b*) together. Each approach discusses both the internal organisation and the external relations of a mode, but make little effort to describe their use, despite the great interest in the *functions* of a mode. For this dissertation, however, the notion of mode has to fulfil both of these needs, that is, (1) to identify the semiotic resources used in the tourist brochures, and (2) to describe how they work toward specific communicative goals in various contexts and how they guide the reader towards the intended interpretation.

To sum up, the definition of a mode needs to be clear and sufficiently fine-grained, and it has to fit the overall theoretical framework. As I already mentioned above, the definition adopted in this dissertation is based on Bateman (2008, 2009*b*, 2011): I will describe the definition in detail in Section 3.3. It is possible to outline some of the benefits of the approach here, as they are relevant to the next topic of discussion, that is, the notion of rank scale. Bateman's definition is compatible with the GeM model (see Chapter 3), which provides a framework for defining analytical units that in turn enable a consistent description of the semiotic modes.

This is a significant advantage, because the definition of analytical units remains a challenge to multimodal analysis, as the following section illustrates.

### 2.3.2 Rank scale

Several theories of multimodality (see e.g. O’Toole 1994, 2011; O’Halloran 2008*b*) propose that language and image possess a hierarchical structure that may be deconstructed into analytical units. The deconstruction of the semiotic modes is also necessary for this dissertation, as the comparison of artefacts requires the notions of structural constituency and hierarchy. In systemic-functional linguistics, the *rank scale* is a theoretical construct used to describe the compositional hierarchy of language, in which the compositional layers of clause, group or phrase, word and morpheme are organised by the relation of “is a part of” (Halliday and Matthiessen 2004, p. 20).

The concept of rank scale was adapted to multimodal analysis early on by Kress and van Leeuwen (1990). A more elaborate application was later presented by O’Toole (1994), who used the concept to deconstruct various multimodal forms of art, such as painting, sculpture and architecture (for a wide range of later applications, see e.g. O’Halloran 1999*b*; Alias 2004; Guo 2004; O’Toole 2004; Ventola 2011). Furthermore, O’Toole combined the rank scale with the metafunctional principle (see Section 2.3.3 for a discussion) into a framework for describing the functions and systems in painting (O’Toole 2011, p. 24). The rank/function matrix has been frequently used to describe the interaction and relations between language and text in various contexts (see e.g. O’Halloran 2005, 2008*b*; Martinec and Salway 2005; Guijarro and Pinar Sanz 2008), although O’Toole has later pointed out that there was no intention to present the rank scales for language and image as corresponding structures with direct mapping in both semiotic resources (Thomas 2009*a*, p. 54).

At this point, it should also be noted that the notion of a rank scale is not uncontested. Zhao (2010*b*), who problematises the application of rank scale in multimodal research, notes that the issue has been extensively debated in multimodal research (cf. Martinec 2005). To put it simply, the issue is whether the segmentation of multimodal data into analytical units advances the analysis or leads to “infinite detail” (cf. Forceville 2007, p. 1236). Zhao also observes that “the flexibility in theoretical conceptualisation ... has led to confusion and low consistency in analytical practices” (2010*b*, p. 254). Moreover, Zhao (2010*b*, pp. 261-262) argues that the relationship between multimodal research and SFL has been so far largely monolateral as opposed to dialogic. In this connection, it is useful to recall the notion of SFL as an exotropic theory presented in Section 2.2.1 and the dialogic principles behind such theories. While the contribution of SFL to multimodal analysis is unarguably significant, it appears that the short-

term, ‘purpose-oriented’ borrowing can have a negative effect on the theoretical development of multimodal research in the long run.

Nevertheless, the work of O’Toole has remained influential in multimodal research, especially in the efforts to determine corresponding structures across semiotic resources and their interaction within the SF-MDA framework (see O’Halloran 2004*b*, 2008*b*). In this approach, the rank scale has proven particularly strong in the analysis of film and television, where it has been used to deconstruct the complex and intensive processes of multimodal meaning-making (cf. Tan 2009; O’Halloran et al. 2010, 2011). Because the rank scale continues to be an important tool for the segmentation of multimodal data and is still being developed (Boeriis and Holzanova 2012), the concept and its applicability to the analysis of print media — in combination with the notion of metafunctions — will be considered again in Section 2.5.2. I will now continue with a description of the metafunctions, which are another central theoretical concept often deployed together with the rank scale.

### 2.3.3 Metafunctions

The concept of *metafunction* is another influential theoretical construct that has been carried over to multimodal research from SFL and social semiotics. The metafunctional principle was introduced in the early work of Halliday (1970/2002*a*, 1973, 1978) to simplify the process of accounting for the form and functions of language. As Halliday points out in an interview by Paul Thibault (1987, p. 607):

[T]he notion of metafunction is simply an attempt to capture this relationship between the internal forms of the language and its use in contexts of social action.

The three metafunctions — the ideational, the interpersonal and the textual — are fulfilled simultaneously in every instance of language use. According to Eggins (2004, p. 2), these metafunctions realise the different but interwoven “strands of meaning”, which O’Halloran (2008*b*, p. 444) describes in the context of multimodal meaning-making as follows:

The metafunctional principle is the principle that semiotic resources simultaneously provide the tools for constructing ideational meaning (i.e. experiential meaning and logical relations) and for enacting social relations (i.e. interpersonal meaning). These metafunctions are enabled through the organisation of discourse, which is the textual metafunction of semiosis.

The quote above illustrates how the concept of metafunctions can be applied to describe the functions of a given semiotic resource. The metafunctions are regarded

as omnipresent, simultaneous and dependent on each other. As a result, no linguistic or multimodal ‘text’, whether written or spoken, can be purely ideational, interpersonal or textual, as the realisation of a text requires contributions of each metafunction. At the same time, each metafunction produces a different type of structural output in language (Halliday 1979/2002*b*, p. 200).<sup>6</sup> Finally, as the metafunctional principle has already received considerable attention in linguistic research (see e.g. Halliday and Hasan 1986; Martin 1991; Jones and Ventola 2008). I will mainly focus on the application of metafunctions in multimodal research.

The works of Kress and van Leeuwen (1996, 2006) and O’Toole (1994, 2011), which have wielded considerable influence on the development of multimodal research, have also extended the metafunctional principle to multimodal analysis. Both Kress and van Leeuwen and O’Toole used the metafunctional principle in their theoretical frameworks to analyse the semiotic construal of various multimodal artefacts or their parts, such as advertisements, photographs, paintings, architecture, sculptures, etc. Because O’Toole focused mostly on art and architecture, I will pay special attention to the work of Kress and van Leeuwen and the particular research stream originating from their work, which are closer to the interests of this dissertation.

In Kress and van Leeuwen (1996, p. 183), the authors sum up the aim of their social semiotic approach to multimodal analysis:

We seek to be able to look at the whole page as an *integrated* text. Our insistence on drawing comparisons between language and visual communication stems from this objective. We seek to break down the disciplinary boundaries between the study of language and the study of images, and we seek, as much as possible, to use compatible language, and compatible terminology in speaking about both, for in actual communication the two and indeed many others come together to form integrated texts.

To achieve the required integration, Kress and van Leeuwen (1996, pp. 40-42) adopted the metafunctional principle to account for the functions that visual communication needs to serve. In addition, the metafunctions provide a framework for contrasting the processes of verbal and visual meaning-making. For instance, Kress and van Leeuwen (1996, p. 49) illustrate how ideational meanings are instantiated using the linguistic system of TRANSITIVITY, which is used to establish

---

<sup>6</sup>For a more detailed discussion of the relation between metafunction and structure, see Martin (1996, 1997). For a discussion of the issue in a multimodal context, see O’Halloran (2008*b*, pp. 447-448).

a relationship between a process and its participants.<sup>7</sup> In image, TRANSITIVITY is realised using vectors: I will present an example later in Section 2.4.2.1.

Although the influence of Kress and van Leeuwen's approach is evident in the wealth of research following the social semiotic tradition, which also speaks for a high descriptive capability and adaptability to describe various types of data (see e.g. the following edited volumes: Baldry 2000*b*; O'Halloran 2004*b*; Ventola et al. 2004; Jones and Ventola 2008; Ventola and Guijarro 2009; Jewitt 2009*c*; Dreyfus et al. 2010; Bednarek and Martin 2010; Baldry and Montagna 2011), there are some concerns about the capability of the metafunctional principle to serve the needs of this dissertation. These concerns are now addressed below.

Let us begin by considering the metafunctional principle as a key heuristic for multimodal analysis. Machin (2009, p. 182) argues that the metafunctions (in combination with the rank scale) allow a more precise description of communication and its multimodal aspects than the approaches that separate the analysis of different semiotic resources and their contributions. As we have seen, the concepts of metafunctionality and rank scale have proven capable of deconstructing the build-up of meaning in multimodal artefacts. While the metafunctions and rank scale may bring us closer to an understanding of how stratified semiotic systems work (cf. Halliday 1979/2002*b*, pp. 196-197), that is, how meanings are structured on the strata of content and expression and the principles that govern these semiotic processes in multimodal contexts (cf. O'Halloran 2008*b*), several open questions about the limitations and capabilities of the metafunctional principle remain.

For instance, Kress and van Leeuwen (2002, p. 345) have discussed the emergence of the semiotic resources within a culture:

It is clear that cultures do not expend the same energy at all times on all the potentially usable semiotic resources: hence some are highly developed and become fully articulated for all the communicative and representational purposes of that group, while others are partially articulated or hardly at all.

What Kress and van Leeuwen suggest is that all semiotic resources are not necessarily similar in terms of the metafunctional structure. This means that although a semiotic resource may possess a metafunctional structure, certain aspects of the organisation may be underdeveloped, especially if contrasted with other semiotic resources at play. Concerns about the extended applicability of metafunctions have also been raised in O'Halloran (2008*b*, p. 451), who points out the need to develop

---

<sup>7</sup>In SFL, the names of systems are commonly written in small capitals. I will follow this convention in this chapter. In later chapters, the small capitals are used to indicate rhetorical relations.



alternative approaches to “cross-functional systems”, such as colour and typography, which operate in the expression stratum. Stöckl (2004, p. 12) also classifies colour as a sub-mode of both language and image, implying that colour is not a fully-fledged semiotic resource (see Figure 2.3). At the same time, colour may have various communicative functions which may be captured using the metafunctional principle, suggesting that in certain contexts, colour has the meaning potential of a full-blown semiotic resource (Kress and van Leeuwen 2002; van Leeuwen 2011).

Given the uncertainty regarding the development of semiotic modes within a culture and their functions, it is necessary to consider whether the metafunctional principle is a methodologically robust starting point for deconstructing the structure of a multimodal artefact. Without a prior knowledge of (1) how the semiotic resources and their functions have developed within a culture, i.e. how much energy has been expended into shaping them, and (2) how these semiotic resources and their configuration may be contrasted to other artefacts and social situations, developing a framework to capture the structure of a multimodal artefact becomes a challenging task. Because the metafunctions are primarily concerned with *meaning*, their coupling with particular kinds of multimodal *structure* remains unclear.

Therefore, it may be suggested that without comprehensive empirical research to identify and describe the underlying systems that realise the meanings we describe using the concept of metafunctions, the metafunctions are too broad for describing complex multimodal phenomena, and they are thus limited to making generic observations about multimodality. This is evident especially in the application of the metafunctional principle to the analysis of layout (see Section 2.4.2.3). For this reason, the metafunctional principle is subjected to a more critical treatment in Section 2.4, which discusses the social semiotic approaches to the analysis of print media.

## 2.4 Social semiotic analyses of print media

This section presents the previous research on print media, which has been conducted within the social semiotic approach to multimodality. Section 2.4.1 begins with the early approaches to print media, followed by an in-depth description and a critical evaluation of the “visual grammar” proposed in Kress and van Leeuwen (1996, 2006) in Section 2.4.2.

### 2.4.1 Early work

Many of the examples discussed in Kress and van Leeuwen (1996) are drawn from print media artefacts. For instance, by contrasting the front pages of two newspapers, *The Sun* and *Frankfurter Allgemeine*, Kress and van Leeuwen (1996, p. 30)

suggest that the visual features of a tabloid newspaper are used to communicate certain ideas:

The reader of *The Sun* is addressed through the medium of the visual (photographs, screamer headlines, colour) with only a small proportion of the page given over to written language. Pleasure, entertainment and immediacy of apprehension determine how ‘reading’ is constructed here. There is usually only one story on the front page: implicitly this is a statement either that the world is a simple, straightforward place, or that, at the very least, it can be made simple by the *The Sun* for its readers.

Without making any statements about *The Sun* and its readership,<sup>8</sup> the quote above outlines certain aspects of print media that need to be described in order to understand how page-based multimodal artefacts work: the role of layout, image, typography and colour, and finally, the written language. However, the quote also reflects the impressionistic side of Kress and van Leeuwen’s work. The verbal and visual conventions of a tabloid newspaper, as opposed to a broadsheet, are seen to reflect intellectual simplicity, although the use of the semiotic resources is more likely to be motivated by the established design conventions (Kostelnick and Hassett 2003, p. 55) and the engaging and inviting function of cover pages (Held 2005, p. 176). Consequently, if we wish to have a firmer analytical grasp of print media, our grip needs to be considerably tighter and we need to be able to make *explicit* statements about both medium and multimodality.

For this reason, I will now proceed to look at the foundational concepts of Kress and van Leeuwen’s theory and discuss their relevance for the current work, in order to find out whether their methods are sufficient for this dissertation.

### **2.4.2 *Visual grammar***

The work of Kress and van Leeuwen (1996) is strongly rooted in the social semiotic tradition, whose principles and key theoretical concepts were introduced in Sections 2.2 and 2.3. To reiterate, Jewitt (2009*a*) outlines the goals of social semiotic analyses of multimodality as follows:

A primary focus of the social semiotic multimodal analysis is on mapping how modal resources are used by people in a given community/social context, in other words sign-making as a social process. The emphasis is on the sign-maker and the situated use of modal resources.

---

<sup>8</sup>See Eggins and Iedema (1997) and Machin and van Leeuwen (2005) on multimodality and target audiences.

As Section 2.3 described, Kress and van Leeuwen apply several of the theoretical concepts developed within systemic-functional linguistics and social semiotics to the multimodal analysis of print media. This section will now look at the application of these concepts to multimodal analysis using the data of this dissertation. The double-page shown in Figure 2.4 is used to exemplify and discuss the kind of analysis enabled by the framework of *visual grammar* proposed in Kress and van Leeuwen (1996, 2006). I begin with a description of TRANSITIVITY, which was briefly introduced in Section 2.3.3.

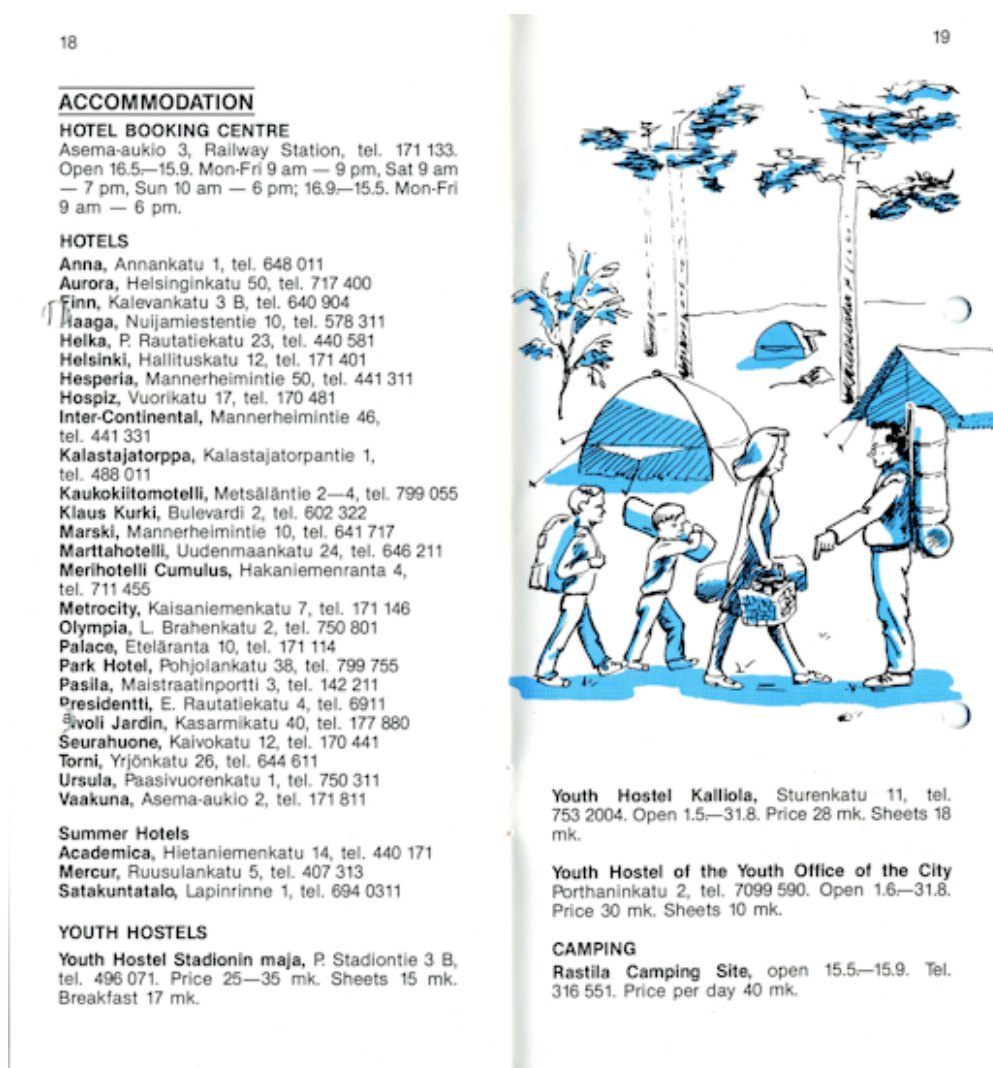


Figure 2.4: A double-page from *Holiday in Helsinki* (1985)

### 2.4.2.1 Transitivity

The previously introduced linguistic system of TRANSITIVITY, which “construes the world of experience into a manageable set of process types” (Halliday and Matthiessen 2004, p. 170), is used by Kress and van Leeuwen to capture the processes of representation in images. In language, a clause is built around a process, with its participants and circumstances in peripheral positions. Visually, a process may be represented using vectors that connect the represented participants (Kress and van Leeuwen 1996, p. 44). The visual narrative shown in the illustration in Figure 2.4 can be used to exemplify this issue. Firstly, the man (*Actor*) on the right performs a *transactional action* (Kress and van Leeuwen 1996, p. 61), as a vector emanates from his hand, which points towards the ground (*Goal*). Both the man and the woman also function as *Reactors*, participating in a *reactional process with sequential bidirectionality* (Kress and van Leeuwen 1996, p. 64), realised by a vector between their eyelines. The *Circumstances* surrounding this action include both *Accompaniments* and a *Setting*, the former realised by the children in a peripheral position, and the latter by the camping grounds and trees, which provide the backdrop for the action.

As the description above shows, the application of TRANSITIVITY analysis to images is useful for a rich description and to some extent, for describing the interaction between language and image (see e.g. Hiippala 2007; Maiorani 2008; Yang 2008; Motta-Roth and Nascimento 2009; Francesconi 2011). However, the use of TRANSITIVITY patterns to model the structure of a multimodal artefact warrants caution for several reasons.

Firstly, multimodal research assumes that meanings arise from the intertwined contribution of multiple semiotic resources. Yet TRANSITIVITY is only one of the systems identified in image: additional systems would have to be defined to complement TRANSITIVITY and to provide a comprehensive view of the entire artefact. Furthermore, a large number of systems would have to be defined in order to capture even the most basic features of print media (see e.g. Lim 2004*a*; O’Halloran and Lim 2009). This results in a rapid loss of analytical focus, because these systems need to be mapped as well to provide a full picture of the artefact’s structure.

Secondly, TRANSITIVITY would have to be complemented by more abstract categories to identify the different kinds of images and their functions, together with their contexts of occurrence (see e.g. Bell and Milic 2002; Machin 2004; Caple 2009*a*). The concept of *coding orientation*, that is, how the images represent their content, could be suggested to fulfil this requirement (Kress and van Leeuwen 1996, pp. 170-171). However, determining the used coding orientation is not a straightforward process, as a closer look at the analysis of images reveals. I will

return to this issue shortly in connection with the system of MODALITY, which is introduced below.

### 2.4.2.2 Modality

MODALITY is the linguistic system responsible for construing the “region of uncertainty” between the positive and negative poles (Halliday and Matthiessen 2004, p. 147). Halliday and Matthiessen (2004, pp. 618-619) present four types of modality: probability, usuality, obligation and inclination. The operation of MODALITY may be highlighted using the following example, which falls into the category of probability:

1. Helsinki *is* the most beautiful city in Finland.
2. Helsinki *may be* the most beautiful city in Finland.
3. Helsinki *is not* the most beautiful city in Finland.

In the examples above, (1) represents the positive pole, while (3) is the negative pole. In (2), the modal operator (*may*) places the statement towards the negative pole (cf. Halliday and Matthiessen 2004, p. 619). To put it simply, MODALITY is concerned with the truth value of a statement.

According to Kress and van Leeuwen (1996, p. 160), MODALITY also operates within the visual domain, where it is concerned with “life-likeness”: abstract drawings are of low modality, whereas high-resolution digital photography is considered more lifelike and therefore to be of high modality. In terms of modality, illustrations such as the one shown in Figure 2.4 are somewhere in between depending on their detail and style, but arguably lower than photographs (see e.g. Figure 3.10). Furthermore, Kress and van Leeuwen argue that modality is not only concerned with representing experience and ideational meanings, but it also has an interpersonal aspect in construing “truths aligning with readers” (1996, p. 160). What is considered lifelike or real thus depends on the context of situation and culture, the producers of the image and the intended audience.

Image (as well as language) can be used to create imaginary representations of situations, which are of appropriate modality in their context of use. In Figure 2.4, the illustration is accompanied by a list of hotels, youth hostels and camping sites. In terms of modality, it is important to consider what kinds of images are likely to occur in this context. For example, the image could be replaced with an image of higher modality, such as a photograph, but in this context, it would be uncommon to find an image of very low modality — such as an abstract diagram — as it would contradict the overall communicative purpose of the artefact (cf. Royce 1998, pp. 39-40).

The degree modality is also dependent on the employed coding orientation, that is, how reality is represented using multiple semiotic resources. Kress and

van Leeuwen (1996, pp. 170-171) distinguish four different coding orientations: technological, sensory, abstract and naturalistic. For instance, the technological coding orientation could represent a touristic landmark using a blueprint and a technical description. In turn, the sensory coding orientation, which aims for aesthetic pleasure, could represent the same landmark using carefully composed and post-processed digital photography. The abstract coding orientation is limited to privileged subjects in artistic, scientific and academic discourse communities and is thus less likely to appear in the tourist brochures. Finally, the naturalistic coding orientation, which represents the world ‘as it is’, is intended to reach a maximally large audience.

For Kress and van Leeuwen (1996, p. 165), the use of colour is also closely linked to the modality of an image and the notion of coding orientation. They suggest that three colour-related factors influence modality:

1. *Colour saturation*, a scale running from full colour saturation to the absence of colour, that is to black and white.
2. *Colour differentiation*, a scale running from a maximally diversified range of colours to monochrome.
3. *Colour modulation*, a scale running from fully modulated colour, with, for example, the use of many different shades of red, to plain, unmodulated colour.

It is reasonable to expect that the data includes images with varying degrees of modality. However, despite the factors presented above, the accurate degree of modality is difficult to measure, therefore rendering the practical analysis of modality unfeasible. As a result, the dissertation will not pursue to develop a specific annotation scheme for visual modality in the tourist brochures, but simply denotes the type of visual element used (see Section 3.2.1).

To conclude, it appears that visual systems such as TRANSITIVITY and MODALITY may be used to tease out certain aspects of images and their structure. If the systems were mapped, they could be used — for instance — to investigate how choices within the systems affect the interpretation of images in the ‘concrete-abstract continuum’ (cf. Messaris 1994, pp. 42-43). The points of interest for this dissertation, however, lie elsewhere (1) in the functions of images in their context of occurrence, that is, the artefact, and (2) in their relation to the other semiotic modes at play (see the large triangle in Figure 2.5). Whether analysing the systems of TRANSITIVITY and MODALITY may help this dissertation achieve its goals is questionable due to the reasons presented below.

To begin with, Forceville (2007, p. 1236) has pointed out the gap between the detailed analyses and the more abstract descriptions and its consequences to theory-building. In short, a bottom-up approach to modelling complex phenomena

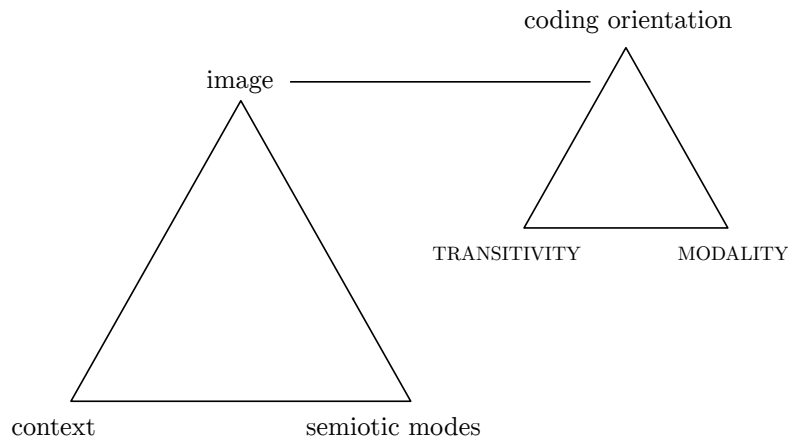


Figure 2.5: Aspects of visual grammar in the analysis of print media

requires robust analytical methods that provide a comprehensive view of the phenomenon at hand. The systems of TRANSITIVITY and MODALITY, which Stöckl (2004, p. 12) defines as “sub-modal” are unable to provide this view, because they are limited to only certain aspects of images (see the small triangle in Figure 2.5).

Moreover, the concept of coding orientation may be far more complex than Kress and van Leeuwen (1996) suggest. Because the production and interpretation of the images using a technological and abstract coding orientation requires scientific and technological training, it might be more appropriate to describe them in terms of a *semiotic mode* than coding orientation. For instance, according to Kostelnick and Hassett (2003, p. 89) technical drawing is a discipline-specific visual code learned through education and enculturation. This brings the process of interpretation to the centre of attention: why would those with a knowledge of technical drawing and its visual code attempt to interpret the drawings in any other way? In this sense, the production and interpretation of technical drawing resembles the operation of a full-blown semiotic mode, whose properties will be described later in Section 3.3. To move forward with the evaluation of the visual grammar, I will now continue with a description of composition.

### 2.4.2.3 Composition

Together with the linguistic systems presented above, the metafunctions are a significant part of Kress and van Leeuwen’s visual grammar. As the influence of the metafunctional principle on multimodal research has already been discussed in Section 2.3, I now focus on the application of the textual metafunction to describe the composition of images and layouts. Kress and van Leeuwen (1996, p. 183)

propose three interrelated systems for describing the composition of an image or a layout:

1. *Information value.* The placement of elements (participants and syntagms that relate them to each other and to the viewer) endows them with the specific informational values attached to the various ‘zones’ of the image: left and right, top and bottom, centre and margin.
2. *Saliency.* The elements (participants and representational and interactive syntagms) are made to attract the viewer’s attention to different degrees, as realised by such factors as placement in the foreground or background, relative size, contrasts in tonal value (or colour), differences in sharpness, etc.
3. *Framing.* The presence or absence of framing devices (realized by elements which create dividing lines, or by actual frame lines) disconnects or connects elements of the image, signifying that they belong or do not belong together in some sense.

These three systems of composition have been the subject of much criticism in recent multimodal research. This criticism is often based on the argument that they greatly simplify a complex phenomenon (see e.g. Bateman et al. 2004; Bateman 2008; Thomas 2009a). Layout composition is arguably one of the most complex areas of multimodal research: it does not only organise the semiotic modes on the page, but may also have implications to visual perception and cognition (Holsanova and Nord 2010; Feng 2011) and the rhetorical organisation of a multimodal artefact (Bateman 2011; Hiippala 2013). Considering that new perspectives to the analysis of layout are still being introduced, it appears that we are only beginning to understand its semiotic potential. However, further advances in this domain are not likely to take place without a considerable amount of empirical research.

A particularly problematic issue in Kress and van Leeuwen (1996) is the overextension of composition to cover *both* layout and image. Research has shown that layouts may take complex and varied forms according to their function (Delin and Bateman 2002; Martinec 2003; O’Halloran and Lim 2009): the same applies to photography (Barthes 1981; Machin 2004; E. et al. 2011). The assumption that the same framework could be used to describe both layout and photography adopts a dangerously simplistic approach towards the description of two fundamentally different phenomena.

Although Kress and van Leeuwen describe the visual grammar of images in great detail, providing an elaborate framework for deconstructing images into participants and for describing the interaction between them, their approach to the interaction of the semiotic resources on a page is rather straightforward. These



concerns were raised at an early stage by Forceville (1999, pp. 171-172), who argues that the framework of Kress and van Leeuwen is far less ‘grammatical’ than it suggests, with very little predictive power and intersubjective validity.

The criteria provided by the three interrelated systems of composition — information value, salience, and framing — are either too general or controversial, and therefore not adequate for describing the composition of language and image on a printed page (Waller 2012, p. 243). This may be explained by an overextension of the metafunctional principle. In the study of language, the ideational and interpersonal metafunctions are part of an established framework with carefully defined systems like TRANSITIVITY and MODALITY, but the textual metafunction does not provide systems that could directly be applied to multimodal analysis (cf. Thomas 2009*a*, p. 45). In order to adapt all three metafunctions to the visual grammar, it appears that the textual metafunction was included without sufficient theoretical consideration and used for describing the most challenging part of the analysis without the necessary tools.

The consequences to multimodal analysis have been severe. Thomas has argued that “these shortcomings are all the more significant because they are reproduced in the literature which has been influenced by Kress and van Leeuwen’s model” (2009*a*, p. 45). His observation reveals a significant gap in the current knowledge of layout and composition in multimodal analysis, whose contribution to meaning-making should not be underestimated. More recently, Kress and van Leeuwen (1996) have also drawn criticism from the field of information design, which has considerable experience in studying layout (cf. Waller 2012).

On the basis of the above discussion, it is clear that the systems of composition cannot be used to describe layout as a part of a model of multimodal structure. This does not, however, rule out the possibility that the previous research may include several useful observations about the nature of different layouts. Consequently, the following section works towards the goal of establishing the possible contribution of previous research to the analysis of layout.

## 2.5 Layout in print media

The previous section established that the analysis of layout, that is, how the semiotic modes combine and interact on a page has been particularly challenging for multimodal research. As I will show below, the previous research may nevertheless inform the work undertaken in this dissertation.

I will first address the work of Royce (1998, 2007) on intersemiotic complementarity in Section 2.5.1, followed by a brief discussion of the stream of research represented by O’Halloran (2008*b*) in Section 2.5.2. I will then consider generic

structures in print media in Section 2.5.3, before finishing with a discussion of covers as a specific page type in Section 2.5.4.

### 2.5.1 Intersemiotic complementarity

Royce (1998) presents one of the first applications of Kress and van Leeuwen's visual grammar to explore the co-occurrence of verbal and visual modes on a page. Royce (1998, p. 27) proposes that the relationship between the modes is characterised by *intersemiotic complementarity*, which is synergistic in nature, as the verbal and visual do not merely co-occur on the page, but they also constantly interact with each other (see also Royce 2007). Furthermore, Royce (1998, pp. 28-29) suggests that the complementarity between the verbal and visual can be approached by combining the visual grammar of Kress and van Leeuwen (1996) with the systemic-functional theories of cohesion. According to Royce, this enables identifying the participants, activities and circumstances present in both language and image, and describing their interrelations using the lexico-semantic sense relations of repetition, synonymy, antonymy, meronymy, hyponymy and collocation (1998, p. 31). In this aspect, he takes a step forward in incorporating the analysis of language and image into a notion of multimodal *cohesion*.

At this point, it is useful to briefly consider the notion of cohesion. Bateman (2008, pp. 145-146) has pointed out that descriptions of multimodal cohesion rarely go beyond superficial observations, unless the analysis is supported by a notion of structure. I propose that cohesion should be viewed as a feature of multimodal semiosis, which arises primarily from meaning-making to consolidate the subject matter realised using multiple semiotic modes. The notion of *coherence*, in turn, is a property of multimodal structure, which ties in with the expectations of the reader and is therefore of higher relevance to this dissertation (see Chapter 4).

Nevertheless, from the perspective of contemporary multimodal research the holistic approach of Royce is remarkably novel in several aspects: the proposed analytical framework is not only limited to language and images, but also accounts for the typographic features of a page. Royce notes that the typographic properties are subject to conventions arising from the particular functionalities of the text in question. For example, the body copy is seen as an orthographic whole, "realised by various typographical conventions" (1998, p. 41). Royce uses similarities in typographic realisation to define structural *units* in the multimodal text: a similar approach can be found in the earlier work of Waller (1987) and later in Bateman (2008, p. 120).

Royce (1998, p. 44) also suggests that the specific page analysed in his study can be divided into four balanced sections and continues by elaborating the relationship between these sections:

In many texts visuals often compete to varying degrees with the verbal aspect of the text for page space. Here however, there is no sense of competition but a clear demarcation and balancing in terms of spatial allocation.

The first statement is somewhat assumptive: it would be more accurate to describe that from a visuo-spatial perspective, language and image occur physically close to each other. Whether they are in competition with each other is difficult to establish and shifts the focus away from the actual use of layout. The second statement can be deemed more accurate: it describes how the modes occupy their own space, in which the semiotic modes realise various functionally motivated elements, such as header, copy, visuals and logo. The question that should follow these two statements should ask why these configurations occur in this artefact, and what their function and structure are in contrast to other configurations in the artefact? This is precisely the question that this dissertation seeks to answer in Chapter 7 by deploying the framework developed here. And as we will see below, in many aspects this contribution resembles the one also envisaged by Royce.

Finally, Royce (1998, p. 46) makes an important statement regarding multimodal analysis of page-based texts and a step towards answering the aforementioned question:

[The WMI text] is an amazingly rich source of complexity in multimodal meanings and *intersemiotic complementarity*, and that there are further areas which would be rich avenues of exploration, and levels of delicacy in analysis which could be developed further.

Underlining the possible analyses on other levels of delicacy, Royce (1998, p. 46) continues:

One area which may be very interesting is how the visual and verbal modes realise *intersemiotic complementarity* at the level of a multimodal text's discourse structure, and even in terms of texts as realisations of a particular generic structures.

Royce's observations regarding future multimodal research are interesting from a contemporary perspective, as the research has often focused on image-text interaction in great detail, while the abstract levels of semiotic organisation have received less attention in multimodal research (see however Lemke 1999, 2000). A full-page advertisement in *The Economist* exemplifies just one page type: within the higher level of organisation of a multimodal artefact, various layouts (and types of intersemiotic complementarity) are likely to be found within the entire magazine.

Instead of looking only at specific, hand-picked examples of image-text interaction, the research needs to be directed towards the larger units formed by the

modes and the relations holding between them. Thus, Royce’s call for expanding the scope of multimodal analysis towards the more abstract levels is on point: it is a prerequisite for a more comprehensive account of multimodal structures in print media. But at the same time, such an account has to be constrained by the notion of structure to retain the capability to compare data (cf. Bateman 2008, p. 145).

In conclusion, the work of Royce (1998) has been mainly followed up in studies of interrelations between semiotic resources (see e.g. Royce 2007; Liu and O’Halloran 2009; Bowcher 2012; Bowcher and Liang 2013). It has also been incorporated into more extensive models of multimodal meaning-making, such as the framework presented in O’Halloran (2008*b*), which I now proceed to discuss.

### 2.5.2 Analysing print media using SF-MDA

O’Halloran (2008*b*, p. 450) presents an elaborate theoretical framework for the analysis of print media, which incorporates several previously introduced theoretical concepts, such as rank scale and metafunctions (see Sections 2.3.2 and 2.3.3) into a framework known as the Systemic Functional-Multimodal Discourse Analysis (hereafter SF-MDA). Following a Hjelmslevian tradition, SF-MDA incorporates two distinct strata: content and expression. Figure 2.6 shows how the rank scales for language and image are positioned in a corresponding relation on the content stratum. The “linguistic, symbolic and visual elements” engage in intersemiosis, which is described using intersemiotic mechanisms (O’Halloran 2008*b*, p. 453). The meanings formed on the content stratum are then realised on the expression stratum by cross-functional systems, such as colour and typography (see Lim 2004*a*).

O’Halloran (2008*b*, pp. 461-469) exemplifies the use of the SF-MDA framework by analysing a printed public service announcement for AIDS awareness. The analysis proceeds from the abstract semiotic levels towards the content stratum, beginning with the generic structure of the advertisement (see Section 2.5.3) by establishing the primary functions of the visual and verbal elements and the layout areas that they occupy. The rank scale for image is used to identify a Scene, which acts as the Locus of Attention in the generic structure, which “embeds the central idea of the advertisement” (Cheong 2004, p. 165). O’Halloran (2008*b*, pp. 464-468) then proceeds with a detailed analysis of the advertisement, which outlines how the meanings emerge as a result of the intersemiotic mechanisms.

O’Halloran convincingly shows that SF-MDA is capable of describing print media advertisements on various levels of analytical delicacy with a high number of systems and functions described in O’Halloran (2008*b*, pp. 457-459) providing the theoretical backdrop. At the same time, the high number of intersecting systems and functions warrants caution, especially from the perspective of theory-building. Despite their high descriptive capacity, the question is whether the systems and

IDEOLOGY	
GENERIC MIX	
REGISTERIAL MIX	
CONTENT Stratum	← INTERSEMIOSIS → Mini-Genres, Items and Components (Linguistic, Visual and/or Other)
	LANGUAGE                      VISUAL IMAGES
	← INTERSEMIOSIS → Discourse Semantics
	Discourse                      Intervisual Relations Work
	← INTERSEMIOSIS → Grammar
	Clause complex                      Scene Clause                                      Episode Word Group / Phrase                      Figure Word    Part
DISPLAY Stratum	← INTERSEMIOSIS → Materiality
	Typography/Graphology and Graphics

Figure 2.6: SF-MDA framework for printed texts: language and visual imagery (Reproduced from O'Halloran 2008*b*, p. 456)

functions have been mapped to a sufficient degree and subjected to empirical scrutiny. The linguistic systems and functions are likely to be more developed in this aspect, whereas the analysis of the visual systems may have been largely exploratory (see e.g. Lim 2004*a*; O'Halloran 2008*b*).

To conclude, I suggested above in Section 2.4.2 that the analysis of visual grammar has to be supported by a more abstract framework, in order to make statements about the generic structure of print media advertisements. However, these generic structures need to be positioned in their context of occurrence — in this case a men's magazine (O'Halloran 2008*b*, p. 469) — which adds another level of analytical delicacy: the multimodal artefact. Several different generic structures are likely to be found within an artefact, whose semiotic properties are shaped by

their function. While remaining within the SF-MDA framework, I now proceed to describe one possible approach to describing these generic structures.

### 2.5.3 Generic structures

Cheong (2004, p. 164) presents a study of five print advertisements. Based on the analysis, Cheong proposes a generic structure for print advertisements: *Lead*  $\hat{}$  (*Display*)  $\hat{}$  *Emblem*  $\hat{}$  (*Announcement*)  $\hat{}$  (*Enhancer*)  $\hat{}$  (*Tag*)  $\hat{}$  (*Call-and-Visit Information*), where  $\hat{}$  indicates ‘followed by’, while parentheses imply optionality. The aforementioned elements of the generic structure allocate the verbal and visual elements into functional clusters, which work together in the print advertisement.

Cheong (2004, pp. 175-176) acknowledges the challenges in describing print media advertisements due to the constant flux in advertisement design. This seems reasonable, as the fluctuation in design conventions has found support in research elsewhere (see e.g. Kostelnick and Hassett 2003). In relation to the data of this dissertation, Molina and Esteban (2006, p. 1051) observe that the tourist brochures are also constantly redesigned. Cheong (2004, pp. 175-176) points out that due to the structural evolution of the advertisements, defining the labels for the required and optional elements is challenging. If this holds true for other multimodal artefacts as well, then we need new approaches to describing the generic structures, which are not based on the principle of linearity (see Section 4.1.1).

As the advertisement for Guess (Cheong 2004, p. 190) shows, advertisements may be minimal in terms of language-image interaction, which leaves the advertisement rhetorically open to interpretation on the level of a page. Instead, the advertisements draw on the meaning potential of the visual to represent the advertised entity. In this case, what could be roughly characterised as the semiotic mode of fashion photography, is responsible for communicating with the reader. This is precisely the area of research to which SF-MDA can contribute to with its high descriptive capability.

Finally, unlike full-page advertisements, the content pages can be far more constrained in their expression, especially in terms of the organisation of language and images in a layout.<sup>9</sup> Although layouts also evolve over time, they may lack the freedom of expression available to print media advertisements. I will return to the topic of advertising in Chapter 6, in which I consider its implications to modelling the artefact structure. The current discussion continues with another page type that may be found in the tourist brochures, that is, the cover page.

---

<sup>9</sup>Legal requirements may also affect layout and design conventions, as Berazhny (2012) and Kostelnick and Hassett (2003, pp. 110-112) have observed.

## 2.5.4 Covers as contact texts

A multimodal artefact may combine multiple page types, of which each may exhibit variation in the multimodal structure according to its function. For example, Held describes magazine covers as “omnipresent contact texts” (2005, p. 173), emphasising the dual function of cover pages as advertisements for the magazine and as a “window” into its contents. According to Held, magazine covers are a “contact-and-advertisement genre”, which “achieve their global meaning on various levels, viz. page composition, layout, visual images and language” (2005, p. 193). Yet attaching the concept of genre to magazine covers should warrant caution, especially as the constraints and the features of the proposed genre are not outlined. A more accurate description would be to suggest that the covers constitute one possible page type in the particular multimodal artefact, which in turn is affected by other constraints that demand the inclusion of this type of page in the artefact.

Held (2005, p. 177) also acknowledges the relationship between the magazine cover and its contents:

... the visual and textual elements are full of cataphoric reference and indices which obviously drag the reader ‘inside’ while simultaneously controlling and supporting his/her navigation.

Here Held relates the magazine cover to the contents, which both co-exist in the same artefact. In terms of layout design, it is reasonable to expect that several layout configurations may be found within a single multimodal artefact. Therefore, the configuration of the deployed semiotic modes needs to be accounted for in each case. The variation exhibited by different page types can then be related to how layout is used to present the content. This, in turn, sets further requirements for understanding how the pages work both individually and together in an artefact: in order to achieve a complete picture of a multimodal artefact and its structure, the analysis needs to be able to take on all page types. Achieving this capability is the main goal of Chapter 3, which is presented shortly after the final remarks.

## 2.6 Concluding remarks

In addition to introducing the field of multimodal research and setting the course of this dissertation, the aim of this chapter was to establish and review the major theories of multimodality and to evaluate how they may be applied in this dissertation. It was established that many of these approaches, such as those of O’Toole (1994, 2011), Kress and van Leeuwen (1996, 2006) and O’Halloran (2008*b*), draw heavily on systemic-functional linguistics and social semiotics for their theoretical concepts and methodology (Jewitt 2009*a*). I will now reiterate the central

challenges in applying these theories to modelling the structure of a multimodal artefact.

Firstly, the shared theoretical concepts for the description of verbal and visual elements, such as rank scale, metafunctions, intersemiotic mechanisms and cross-functional systems, provide the approaches with a high descriptive capability and adaptability to various data, which is reflected in the wealth of subsequent research. Yet it needs to be understood that the primary concern of these approaches is meaning, not structure. For this reason, what these approaches gain in the high degree of integration in the analysis, they lose in the capability to segment the data into analytical units for the analysis of structure. Without corresponding analytical units, it is questionable whether the analyses are comparable with each other.

Secondly, a clearly defined notion of structure is necessary for any investigation that seeks to compare multimodal phenomena. The evaluated approaches largely lack this capacity, which is reflected in the absence of large-scale investigations driven by corpus analyses, although their importance has been frequently highlighted (see e.g. Baldry 2007; Parodi 2010; Bateman 2012). Moreover, the absence of structure manifests itself as analytical problems when applying theoretical concepts such as rank scale on data, while moving *from one rank to another* on the rank scale, that is, describing how elements combine into larger wholes (cf. Zhao 2010*b*).

To conclude, given the state-of-the-art in multimodal research, a trade-off between descriptive and comparative capability is inevitable. In the following chapter, I investigate what might be gained by investing into a structure-driven, comparative approach to multimodality.



# Chapter 3

## An empirical approach to multimodality

The third chapter of this dissertation turns towards an empirical approach to multimodal analysis. By discussing a particular model of multimodal document structure — the *Genre and Multimodality* (GeM) model — the chapter seeks to show that an empirical, corpus-driven approach is currently the most viable option for developing a model of a multimodal artefact. Moreover, the chapter also emphasises the emerging notions of semiotic mode, which are applicable to a wide range of contexts, and possess an increased analytical capability and explanatory power due to theoretical refinements.

The chapter begins with an introductory description and an overview of the GeM model in Section 3.1, before proceeding to discuss its analytical layers in greater detail in Section 3.2. This discussion is then followed by a closer look at the semiotic modes in Section 3.3. Finally, Section 3.4 is dedicated to a specific semiotic mode, *page-flow*, which takes advantage of the layout space to organise the content. At this point, it is important to note that the actual application of the GeM model will not be described here: this issue will be taken up in Chapter 5. This chapter also provides the basis for a take on the notion of genre in Chapter 4, which seeks to explain the multimodal structures observed in the data.

### 3.1 The *Genre and Multimodality* (GeM) model

The *Genre and Multimodality* model (hereafter GeM; see Bateman 2008) is a model for describing the structure of a multimodal artefact. As the name of the model suggests, genre is a foundational notion within the model. Genre provides the means (1) to compare artefacts, and (2) to theorise how the artefacts deploy the semiotic modes (Bateman 2008, pp. 9-10). In this chapter, I argue that the GeM

model is currently the strongest candidate for providing the analytical tools needed for modelling the structure of the tourist brochure. The GeM model advocates an empirical approach to studying multimodal artefacts, that is, identifying and analysing multimodal phenomena, formulating hypotheses, checking them against corpora and feeding the results back into the theory (Bateman 2008, pp. 14-15). This is a sensible point of departure, given our lack of experience in studying multimodality on the scale of an artefact.

Bateman (2008, p. 2) outlines the aims of the GeM model as follows:

The model proposed is intended to allow researchers and practitioners alike to attack any example of a multimodal document with a single set of tools that can provide *reproducible*, and therefore *evaluable*, analyses of what is involved in the multiplication of meanings discovered.<sup>1</sup>

The key concepts in the quote above are ‘reproducible’ and ‘evaluable’. The GeM model provides a set of analytical tools and a consistent method, which allows the dissertation to describe and track the changes in the structure of the tourist brochures. This, in turn, allows the identification of reoccurring phenomena, which may yield valuable clues about the structure of the artefact. In this way, the model enables this dissertation to take on the data, which spans over four decades and can thus be expected to include variation in their use of language, image and their combinations in a layout.

This obviously anticipates the following question: why is it necessary to account for *change* in this dissertation? Adding to the criticism of a static perspective to semiosis, which was discussed in Section 2.1.2, Kress (1998, p. 74) has argued that:

[T]he semiotic landscape is changing in fundamental ways, and this change relates to other changes in social, cultural and economic and technological domains ... [a] semiotic theory which does not have an account of change at its core is simply inadequate and implausible in the present period.

The perspective of this dissertation to the issue of change is shown in Figure 3.1, which presents three interrelated factors that are likely to shape artefact structure over time.

Firstly, *the semiotic modes and media* are known to change over time (see e.g. Twyman 1986; Baldry 2000*a*; Labov 2001; Phillips and McQuarrie 2002; Chambers et al. 2002; Kostelnick and Hassett 2003; O’Halloran 2009*a*): these constitute the first factor of change. I will discuss the semiotic modes in Section 3.3, followed by an extensive discussion of the brochure as a medium in Chapter 6.

---

<sup>1</sup>My emphasis.

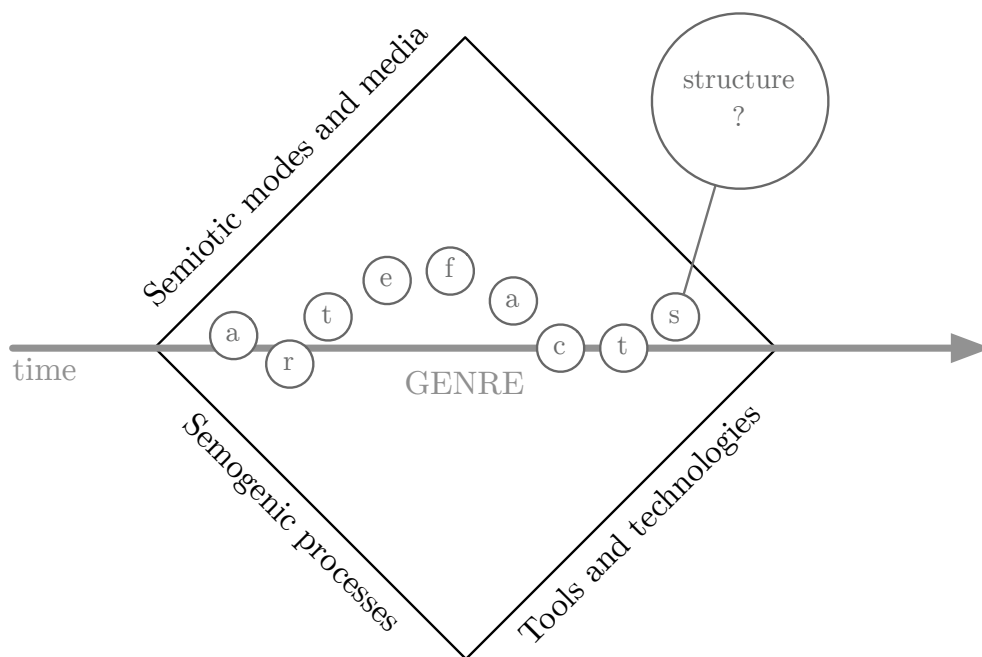


Figure 3.1: The evolution of multimodal artefacts and genres

Changes in the semiotic modes result from the second factor, the *semogenic processes*, which account for the development of meaning-making in both individuals and groups (cf. Lemke 2000; Rose 2006). As a result, both “sign and symbol systems” (Ivarsson et al. 2009, pp. 202-203) and the way multimodal artefacts are used (Allen et al. 1999, p. 7) are subjected to the same dynamics of change as the semiotic modes and media. A discussion of these processes will be presented in Section 4.3.2.

Finally, *tools and technology*, such as printing, desktop publishing and digital photography create space for the expansion of the semogenic processes by providing new means of manipulating the semiotic modes and media. The available tools and technology have already been noted for their significant role in shaping the semiotic landscape (O’Halloran 2009a, p. 99). Their impact is briefly discussed in Section 3.5.1 and more extensively in Section 6.3.2.

To conclude, this section has attempted to outline the factors that shape the multimodal artefacts, which need to be accounted for when describing their structure. I also suggested that incorporating the notion of change into the framework is essential, because the semiotic modes and the tools that are used to realise and produce the multimodal artefacts are constantly transformed. This is kept in mind, as the dissertation describes the GeM model and evaluates its applica-

bility to modelling artefact structure. To provide an in-depth perspective to the framework, I now proceed to discuss each analytical layer of the GeM model.

## 3.2 The analytical layers of the GeM model

The GeM model has four analytical layers: the base, layout, rhetorical and navigation layers. I will now discuss each layer in detail and describe how they contribute to this dissertation, while also highlighting the possible challenges in their application. The discussion begins with a description of the base layer, which is used to segment the data into analytical units.

### 3.2.1 Base layer

It should be noted that few, if any, of the multimodal frameworks attempt to strictly define a minimal unit of analysis (cf. Jewitt 2009*a*). Instead, the analytical focus may be increased or decreased as required (see e.g. Kress and van Leeuwen 1996, 2006; O’Halloran 2008*b*; Boeriis and Holsanova 2012). As a result, any part of the studied artefact may be picked up for analysis: there are no constraints on the analyst’s treatment of the artefact. This is precisely what invites the “infinite detail” in multimodal analysis of artefacts, which has raised concerns in previous research (cf. Forceville 2007, p. 1236).

The GeM model, however, draws on the notion of structure for its analytical capability and thus requires a minimal unit of analysis. This is the task of the base layer, which Bateman (2008, p. 110) describes as follows:

The purpose of the base layer is to identify the minimal elements which can serve as the common denominator for interpretative and textual elements as well as for layout elements in any analysis of a page or document. Everything which can be seen on each page of an analysed document should *be assigned to some base unit*.

In short, the base layer provides the other analytical layers with a comprehensive set of analytical units, which are defined according to a set of criteria. These criteria are circumscribed in a list of *Recognised Base Units* (hereafter RBUs), which provides guidelines for identifying the base units.

Table 3.1 lists the RBUs identified in the GeM model, which enable a comprehensive analysis of an artefact, page by page, without resorting to intuitive and arbitrary segmentation. However, the tourist brochures may also contain certain problematic image-text combinations, which embed verbal base units into visual base units, such as maps. A problem may occur when annotating maps, as the GeM model specifies that “text in photos, diagrams, drawings” (Bateman 2008,

Table 3.1: A list of base units

<b>Recognised Base Units</b>			
sentences	headings	titles	headlines
icons	table cells	list items	list labels
footnote label	items in a menu	page numbers	running heads
emphasised text	floating text	<b>maps</b>	

- sentence fragments initiating a list
- footnotes (*without footnote label*)
- photos, drawings, diagrams, figures (*without caption*)
- captions of photos, drawings, diagrams, tables
- text in photos, drawings, diagrams
- horizontal or vertical lines which function as delimiter between columns or rows
- lines, arrows, polylines which connect other units

(Source: Bateman 2008, p. 111)

p. 111) are to be recognised as base units. While this rule will be maintained in this dissertation, maps will be excluded from the rule as they operate under their own conventions defined by the field of cartography (see e.g. MacEachren 1995).

Due to the complexity of cartography, maps would benefit little from multimodal analysis.<sup>2</sup> Consequently, maps will be annotated as independent base units. In other cases, the rule of identifying text-in-image as a base unit will be maintained. The changes are reflected in the amended list of RBUs in Table 3.1.

As for language, the minimal unit of analysis identified in the base layer is an orthographic sentence. What this means is that no unit embedded within an orthographic sentence or a sentence fragment will be picked up for analysis in the layout and rhetorical layers. This level of analytical granularity is regarded as sufficient for this dissertation for several reasons. Firstly, linguistic and multimodal features of tourism discourse have been extensively studied in previous research (see e.g. Guijarro and Hernandez 2001; Bonelli and Manca 2002; Kaltenbacher 2006, 2007; Hiippala 2007, 2012a; Kvåle 2010; Thurlow and Jaworski 2010; Francesconi 2011). It should be noted, however, that the open architecture of the GeM model could be extended to include grammatical analysis, if it would support the goals of this dissertation (Bateman 2008, p. 112). Retaining the orthographic sentence as the

---

<sup>2</sup>It should be noted that multimodal discourse analysis can contribute to some analyses, particularly in cases of specialised maps, when analysed in their context of occurrence (cf. Leander and Vasudevan 2009).

minimal unit of analysis enables the dissertation to focus on the multimodal structure of the artefact, while also keeping its linguistic structure under analytical control.

The task of modelling artefact structure requires maintaining a sharp analytical focus due to the structurally heterogeneous data (see e.g. Hiippala 2012*b,a*), which does not allow an in-depth exploration of specific grammatical phenomena. For the same reason, the segmentation of visual elements, such as photographs and illustrations, into analytical units smaller than the defined RBUs is not within the scope of this dissertation (see also Section 3.3.2). In this way, the study can avoid the problem of losing focus due to the infinite detail in analysis (Forceville 2007, p. 1236). This does not, however, imply that the dissertation adopts a simple and straightforward approach to multimodal analysis. In contrast, the complexity of the multimodal artefact becomes evident with the combined application of the GeM model's analytical layers.

### 3.2.2 Layout layer

The data can include different layout designs, as the tourist brochures have proven highly heterogeneous in terms of form, size and content (Molina and Esteban 2006, p. 1051). These properties arise from the factors affecting the artefact structure, which were introduced earlier in Figure 3.1. And for this reason, the method used for layout analysis has to be able to account for all types of layouts.

In the GeM model, the layout is described by using three interrelated components: (1) layout structure, (2) area model and (3) realisation information. Together, these three components provide information on the hierarchical organisation of the content in the layout, its use of two-dimensional space, and the graphic and typographic features of the base units. Each component will be now described separately, beginning with the layout structure, continuing with the area model, and concluding with the realisation information.

#### 3.2.2.1 Layout structure

The base units often form larger units. For example, a header can be followed by a paragraph consisting of several sentences. According to the RBU definitions given in Table 3.1, the header and each sentence would be annotated as independent base units. Similarly, an introductory sentence followed by a list could serve as another example. These types of hierarchical structures may be visualised using tree diagrams, as shown in Figure 3.2 (for another example, see Bateman 2008, p. 123).

Figure 3.2 shows a hierarchical organisation of the layout units, hereafter referred to as the *layout structure*. As a part of the layout layer, the layout structure

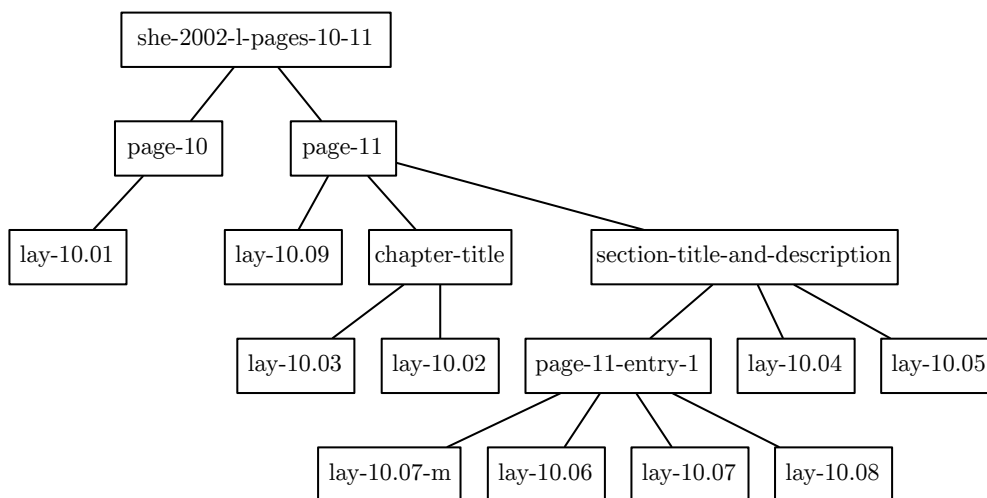


Figure 3.2: An example of a hierarchical layout structure

describes how the content is organised. The individual layout units come together under a parent node — a layout chunk — to form larger units, such as texts consisting of multiple paragraphs in the case of **page-11-entry-1** in Figure 3.2. Above this chunk, we can find additional parent nodes as we move up the layout structure hierarchy: the section, the page, and finally, the entire double-page. For solving the layout structure, the GeM model relies on several methods: I will now describe them below.

Bateman (2008, p. 122) proposes two methods for grouping the base units into layout units. The first method, based on Reichenberger et al. (1996, p. 5), transforms the layout into a digital image and reduces its resolution, which makes the content blend into each other (see Figure 3.3). The resolution reduction can be used to group the layout units together, based on the “visual coherence” of the elements that remain visible after the resolution is reduced. Although this method may be useful for solving the layout structure of an artefact, the method is also time-consuming. Because the second method presented below is more efficient, the method of resolution reduction is reserved for particularly challenging instances encountered in the data.

The second method for determining layout structure is based on the realisation information — a component of the GeM layout layer — which will be described in greater detail in Section 3.2.2.3. This method uses the typographic and graphic realisation information to group the base units into layout units according to their realisational features. For example, if the sentences that form a paragraph share the same typographic features, a strong argument can be made for grouping them



Figure 3.3: *Helsinki's Four Tourist Islands (1988)* with a resolution reduction from 300 DPI to 5 DPI

together into a layout unit in the layout structure. This method is deemed sufficient for most cases in this dissertation.

Finally, it is necessary to consider the layout structure in relation to other types of structure. Reichenberger et al. have proposed that “the strength of visual coherence is proportional to the strength of rhetorical coherence” (1996, p. 5). They thus suggest that the layout structure is not a simple hierarchical organisation. Instead, the layout is also motivated by the needs of rhetoric, that is, how the content is presented to the reader. Consequently, what exactly drives the formation of the layout units is of high interest for this dissertation and multimodal research in general. For instance, the spatially proximal elements in layout have been described as “clusters” (Baldry and Thibault 2005, p. 31), “image-text-complexes” (Kvåle 2010) and as a *Gesamtkunstwerk* (Matthiessen 2007, p. 25). What we need are the analytical tools to take these units apart.

Given the fact that our knowledge of the ‘internal organisation’ of the layout and content is still relatively limited, their analysis should warrant caution. The layout structure offers the means to observe how the base units — the actual content — are organised in the artefacts. While the linear structure of unfolding text and its realisation on the actual physical layout are reasonably predictable (see the discussion on text-flow in Section 3.4.1), the principles that govern the structure of multiple semiotic modes are not yet known to such extent. As I will show later in Chapter 8, considerable advances in understanding the operation of multiple semiotic modes may be achieved by studying the hierarchical layout



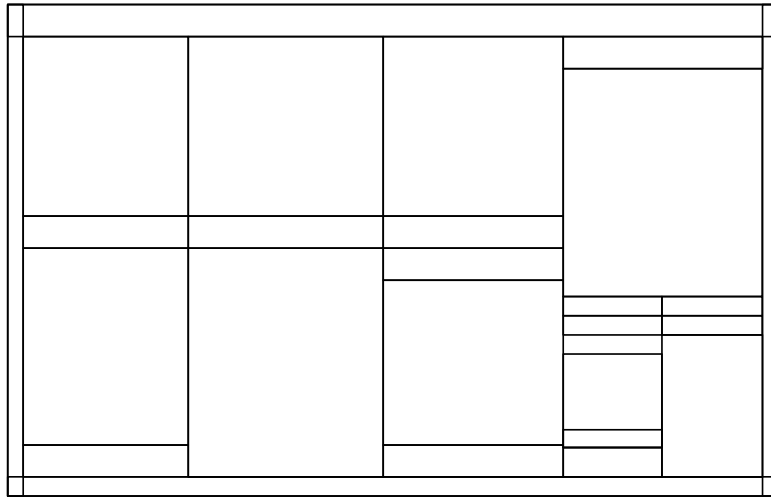
structure and the rhetorical structure in connection with each other. Next, I will move to describe a representation of the physical layout using the area model.

### 3.2.2.2 Area model

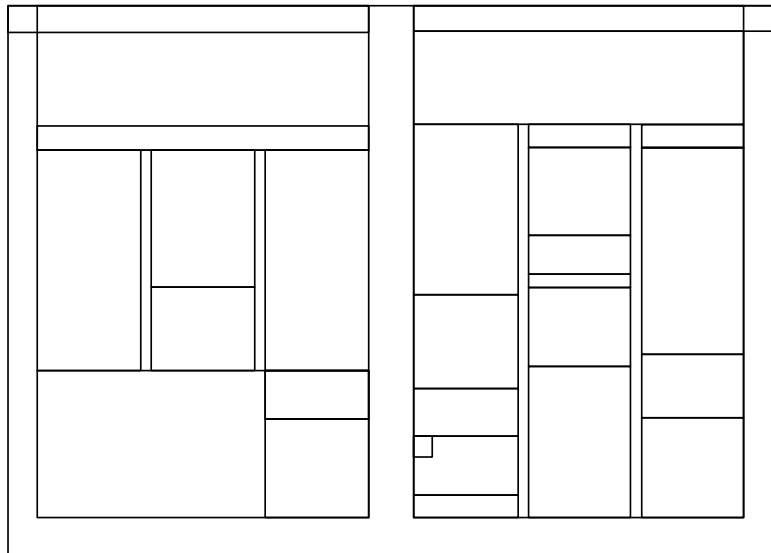
In the GeM model, the area model is responsible for providing a representation of the physical layout of the artefact. Elsewhere, layout has been described from the perspectives of composition (Kress and van Leeuwen 1998, 1996, 2006; Royce 1998, 2007), generic structure potential (Cheong 2004) and cognition (Feng 2011). The area model, however, does not assign any particular functions to layout and its parts, but instead provides information on the spatial organisation of the content using a “typographic” or “baseline” grid, which is a well-established design tool in document, book and graphic design (Williamson 1986).

The primary task of the area model is to provide location information on the layout structure, which is important to this dissertation for several reasons. Firstly, the area model has been used to describe various layouts in both print and digital media (Bateman et al. 2000, 2007) and may therefore be able to meet the challenges presented by the heterogeneous data (see Section 3.2.2). Secondly, the area model does not assign any predefined functions to specific layout areas, which has been a frequently criticised aspect of the social semiotic approaches to layout (see Forceville 1999; Knox 2007, pp. 37-38; Bateman 2008, pp. 40-50, Thomas 2009*a*, pp. 45-46 and Waller 2012). Therefore, the area model provides this dissertation with a neutral starting point for the analysis of layout. In connection with the other analytical layers, the area model contributes particularly to what may be termed *cross-layer analyses* within the GeM model, especially between the layout and rhetorical layers, which may describe together how the semiotic modes are configured for particular kinds of communicative work.

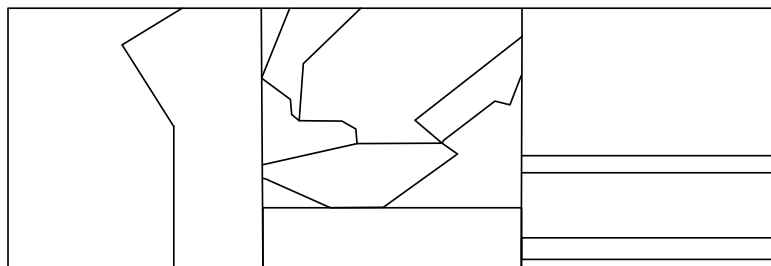
The empirical analysis of layout is important, because relatively little is known about the mechanisms that govern the use of two-dimensional space in multimodal meaning-making, despite extensive theoretical discussions of the issue (see e.g. Iedema 2003; Lim 2004*a*). At first sight, a grid-based area model may seem simple, but when language and images are ‘poured’ into the model, it becomes obvious that a multimodal artefact is a product of complex and interrelated choices. The complexity of the artefact becomes evident when the analytical frameworks are applied to the instances of language and image to analyse their interrelations. Furthermore, the currently available theoretical frameworks are barely capable of scratching the surface, let alone provide a comprehensive picture of the meaning-making processes in a multimodal artefact. The only way to remedy the situation is to observe data, build theories, apply them, and to feed the findings back into the theory.



(a) *Helsinki's Four Tourist Islands (1988)*



(b) *Helsinki Your Way (2006)*



(c) *Bohemian Nordic Oddity (2006)*

Figure 3.4: Three GeM area models

A sufficiently robust theory of multimodality should allow us to make suggestions about the mechanisms of multimodal meaning-making, and particularly about the use of layout space and organisation, which may have far-reaching consequences within the fields of pedagogy, or more generally, in any field that involves the transfer of knowledge (see e.g. Hegarty and Just 1993; Hannus and Hyönä 1999; Hiippala 2012*c*). As Holsanova and Nord (2010, p. 83) have pointed out, it may be suggested that the users of a multimodal artefact draw on previous experiences, which are reinforced by the artefact that is currently interacted with:

[T]he user recognizes functional patterns and principles behind the structure, knows where to look for specific things, how to find entry points and possible reading paths, how to recognize information hierarchies, etc. Thus, the structure of the media offers the reader certain directions, suggests meaningful units, shows possible ways of exploration and guides the reader towards interesting or promising items.

What the theory of multimodality needs is the ability to describe how these patterns and principles emerge and organise themselves, which is also one of the main contributions of this dissertation. This, however, cannot be achieved using *ad hoc* methods, but through careful analysis of multimodal phenomena, such as the configuration of the two-dimensional space of a layout.

Let us now look at several examples of area models in Figure 3.4 to illustrate how the area model works. Initially, the intersecting horizontal, vertical and diagonal lines form a representation of the baseline grid, which has been used to aid the placement of text and image in the layout. As Bateman (2008, p. 80) points out, the grid lines form spaces into which the content may be poured. The organisation of content in the two-dimensional space is not a random process, but subject to conventions (Kostelnick and Hassett 2003, pp. 53-54) and constraints (Bateman and Henschel 2007, p. 31), which may be uncovered using the area model. As Figure 3.4 indicates, the grid is a highly flexible design tool. Therefore, the dissertation has to establish the configurations that help the reader to identify the artefact as a tourist brochure (cf. Holsanova and Nord 2010).

In principle, the description of a layout using the GeM model may be deemed successful, if each layout unit identified in the analysis may be assigned with precise location information. Based on this information, the area model may be used to study the deployment and configuration of the semiotic modes in the layout. For example, it is possible to study whether rhetorically connected verbal and visual elements typically occur within nested or adjacent layout areas. As it was previously suggested, such in-depth observations become possible only when the output of multiple analytical layers are combined (cf. Hiippala 2013). Thus, the following section adds yet another aspect to the analysis of layout: the realisation information, which describes the typographic and graphic features of the content.

### 3.2.2.3 Realisation information

The realisation information describes the typographic and graphic features of the base units. In short, each RBU is described in terms of its typographic or graphic features according to predefined criteria, such as font family, size, weight, colour, style, etc. (see Bateman 2008, p. 120). In addition to providing the criteria for the layout structure (see Section 3.2.2.1), the realisation information allows a perspective into the typographic and graphic aspects of a multimodal artefact and their contribution to multimodal meaning-making. Within the GeM framework, these issues have been previously explored by Delin and Bateman (2002) and Thomas (2009*b*).

Essentially, the realisation information gathers information on what have been termed as “cross-functional systems” in O’Halloran (2008*b*, p. 451) and “sub-modes” in Stöckl (2004, p. 12), which operate on the expression plane (Lim 2004*a,b*). The GeM model does not, however, model graphic and typographic choices as independent systems, but incorporates them directly into the layout layer. The realisation information thus allows the observation of typographic and graphic phenomena in relation to other analytical layers, thereby making it possible to study how the typographic and graphic choices are motivated functionally (cf. Waller 1987).

This leaves us with the more challenging domain of the realisation information, that is, the classification of images. The tourist brochures use images to a great extent, as they contribute significantly to the mental image of a destination (Scarles 2004; Molina and Esteban 2006; Jokela 2011). As Machin (2004, p. 329) has pointed out, in the age of highly decontextualised stock photography, tourism photography still retains a high degree of geographical contextualisation (see also Garrod 2009). For example, a landmark or a sight may be described *visually* using a photograph, an illustration, or in some cases a map, if the reader is being guided through the site (see Figure 3.5 in Section 3.2.4 for an example). The aforementioned challenges emerge as soon as we attempt to categorise and classify the graphic elements in a multimodal artefact.

Bateman argues that “no empirically motivated set of properties” (2008, p. 121) for describing the graphic elements in multimodal artefacts currently exists, suggesting that the ‘visual grammar’ described in Section 2.4.2 warrants caution because it lacks empirical verification. To draw on an analogy from multimodal film analysis, Tseng and Bateman (2012, p. 93) propose that any descriptive method should (1) bridge the gap between detailed and abstract analyses, and (2) “operate without specific commitments drawn from the individual film under analysis”, thus enhancing the model’s capability to compare data. The same principles should apply to the analysis of static multimodal artefacts. The proposal of Bateman (2008, p. 121) for a division of the graphic elements into photographs, naturalistic

drawings, line drawings and diagrams is reasonable, as it increases the comparative capability of this dissertation by providing an abstraction of the graphic elements. For current purposes, the maps are also added to the list.

For language, the realisation information allows the GeM model to describe the typographic features of verbal elements with relative accuracy (Bateman 2008, pp. 117-118), because previous research has largely established the concepts necessary for describing typography (see e.g. Waller 1987, 1990). Therefore, typographic variation and its functions in context are of higher interest than the connotative meanings of typefaces (cf. Stöckl 2005; van Leeuwen 2005*d*, 2006), because the dissertation needs to know if and how typography contributes to the structure of the artefacts. Furthermore, in addition to grouping the base units into layout units, the realisation information can be used to trace the specific functions of typography via cross-layer analyses, as Delin and Bateman (2002) have shown. To take us further towards these cross-layer analyses, I will now move to discuss the rhetorical layer, which adds yet another aspect to the description of multimodal artefacts.

### 3.2.3 Rhetorical layer

As a part of the GeM model, the rhetorical layer aims to “identify the particular functional contributions made by the elements of a document to the intended communicative purposes of that document as a whole” (Bateman 2008, p. 144), that is, how the verbal and visual elements combine into a meaningful whole in a layout. The rhetorical layer draws heavily on Rhetorical Structure Theory (hereafter RST), which is a theory of text structure and organisation developed by Mann and Thompson (1987, 1988). Although Chapter 2 presented criticism of the application of linguistic concepts to multimodal analysis, particularly in connection with the metafunctions, these issues were largely related to the overextension of linguistic concepts. RST, in turn, contributes directly to only one aspect of the GeM model, and its analytic capability emerges largely in connection with the other layers of description.

Since its introduction, RST has generated interest as a theory of text organisation and *coherence* within several fields of study (see Taboada and Mann 2006*a*). Mann et al. (1992, p. 43) make the following hypothesis about text organisation, which explains the principles behind RST:

Texts consist of functionally significant parts; the parts are elements of patterns in which parts are combined to create larger parts and whole texts.

What RST does, then, is this: the theory addresses the relations that hold between sequential parts of a text, which combine into larger parts and entire texts

(Taboada and Mann 2006*b*, p. 425). These patterns have been theorised in SFL using the concept of metaredundancy, which postulates the existence of redundant patterns across the multiple strata of language (Martin 1991, 1999). It is worth mentioning here that the possible points of contact between SFL and RST have also been explored in the work of Mann and Matthiessen (1990) and Bateman and Rondhuis (1997). While the exploration of this issue does not fall within the scope of this dissertation, it is definitely an issue that warrants further investigation, especially in the field of SFL-inspired multimodal research.

For current purposes, the goals of both RST and multimodal analysis need to be compared. Theories of multimodality put forward that the text and its functionality are also shaped by the contribution of non-linguistic semiotic modes (see e.g. Kaltenbacher 2004; Jewitt 2009*a*), which together provide — to paraphrase Stöckl (2004) — the ‘building blocks’ of a multimodal artefact. In this connection, it should be noted that RST was used to describe the structure and coherence in multimodal documents — that is, how the parts of their structure relate to each other — at an early stage of multimodal research (see André and Rist 1995; Reichenberger et al. 1996). Outside the GeM framework, Taboada and Habel (2013) have recently explored the application of RST in the description of multimodal artefacts.

Bateman (2008, pp. 144-145) argues that the current models of image-text interaction in a multimodal artefact have been limited in their scope, thus neglecting the description of how the larger parts of the artefact work together (see e.g. Martinec and Salway 2005; Kong 2006). As a solution to the challenge presented by the description of the overall structure of a multimodal artefact, Bateman proposes an application of RST. Recently, the application of RST has been criticised by Martinec (2013), who argues that RST is not semiotic but a cognitive model, because it cannot explicate the produced analyses, that is, the resulting discourse interpretations.

Martinec (2013, p. 150), however, fails to account for the fact that discourse semantics are arguably semiotic by nature. They are socially shared conceptions of discourse interpretation that have arisen over time as a result of interaction (for discourse semantics, see Section 3.3.3). And to uncover the signalling of these discourse semantics, many other analytical tools may be brought in to complement the performed analyses (cf. e.g. Asher and Lascarides 2003; Bateman 2011; Bateman and Schmidt 2012).

Considering the work undertaken here, it is obvious that the dissertation requires an analytical reach that covers the entire artefact. Therefore, the following discussion focuses on “classical RST” (Taboada and Mann 2006*b*, p. 426) in Section 3.2.3.1, and the application of RST as a part of the GeM model in Section 3.2.3.2.

### 3.2.3.1 Classical RST

According to Taboada and Mann (2006*b*, p. 426), classical RST, as it is generally understood, refers to the initial work of Mann and Thompson (1987, 1988), which defined a basic set of 24 rhetorical relations, which was later expanded to 30 relations in Mann (2005). This section presents the essential aspects of RST and its method in the following points.

Firstly, classical RST does not define a minimal unit of analysis, but as Mann and Thompson (1988, p. 248) point out that in their analysis, the defined units are essentially *clauses*. However, as Section 3.2.1 established, the minimal units of analysis in this dissertation are orthographic sentences and sentence fragments — the analysis does not extend below to the rank of a clause.

Secondly, classical RST defines relations between two non-overlapping text spans, referred to as *nucleus* and *satellite* (Mann and Thompson 1987, p. 4). This means the text spans take on certain roles in a rhetorical relation; the nucleus carries the basic information, whereas the satellite contributes additional information (Taboada and Mann 2006*b*, pp. 426-427). However, in certain rhetorical relations, both text spans may act as the nucleus, resulting in a relation with multiple nuclei. To account for alternative nucleus-satellite configurations, classical RST uses five schemas to define the structural relations between text spans (Mann and Thompson 1988, p. 247). The implications of different schema to GeM RST are discussed in the following section.

Thirdly, RST includes definitions for the rhetorical relations used in the model, which describe the relations holding between the text spans. A relation definition is made up of four fields (Mann and Thompson 1987, p. 245), which are exemplified here using the relationship of PURPOSE<sup>3</sup> (Mann and Thompson 1987, p. 276):

1. *Constraints on the nucleus (N)*: presents an activity.
2. *Constraints on the satellite (S)*: presents a situation which is unrealised.
3. *Constraints on the combination of nucleus and satellite*: S presents a situation to be realised using the activity in N.
4. *The effect*: reader recognises that the activity in N is initiated in order to realise S.

A discussion of each relation defined in Mann and Thompson (1987, 1988) and Mann (2005) is not possible as a part of this thesis, owing to the focus and limited space (for the definitions, see Appendix B). However, the basic principles of RST analysis should be considered here: the analyst evaluates the intentions of

---

<sup>3</sup>The convention of typing the relation names in small capitals, which is common in RST, is also followed in this dissertation. This is not to be confused with the earlier use of small capitals for systems in SFL.

the writer and the desired effect of the chosen rhetorical relation on the reader. As Mann and Thompson (1988, p. 246) acknowledge, such judgements are uncertain, and therefore the statements regarding rhetorical relations are *plausibility judgements*. Although the analytical reliability of classical RST has not been extensively studied, Taboada and Mann (2006*b*, p. 444) report on high analytical consistencies and agreement in several studies that involve modifications of RST and multiple analysts.

To conclude, RST appears to be a formidable theory of text organisation and coherence, which has been applied across a number of different fields of study (for an overview, see Taboada and Mann 2006*a*). Thus the question is, how efficiently RST may be applied to describe the organisation of multimodal artefacts? Several challenges emerge after a brief consideration. For instance, organising principles of the individual semiotic modes are different. Whereas language relies on its sequential organisation, multiple semiotic modes often take advantage of the available two-dimensional space (see Bateman 2009*b*, 2011). In addition, the nuclearity in image-text relations seems to be highly context-dependent (see e.g. Martinec and Salway 2005; Caple and Bednarek 2010) and ambiguous (Taboada and Habel 2013, p. 85). For these reasons, developing a consistent schema for image-text relations appears difficult. These issues and other aspects of GeM RST are considered in greater detail below.

### 3.2.3.2 Extended GeM RST

Bateman (2008, pp. 158-159) highlights several challenges related to GeM RST, which are described in the following points.

The first challenge of extending RST to multimodal analysis arises from its original use for text analysis. As Bateman points out, “conventional RST builds on the sequentiality of text segments” (2008, p. 157). As I argued above, all multimodal artefacts are not organised sequentially. There have been proposals of linear descriptions of multimodality in the form of potential “reading paths” (van Leeuwen 1993, 2005*b*), but because these semiotic models of reader behaviour remain underdeveloped, the application of the reading paths in this dissertation has to be ruled out (see e.g. Holsanova and Holmqvist 2006; Hiippala 2012*c*). Therefore, alternative means of constraining the application of RST schema need to be defined.

In order to constrain the application of RST, Bateman (2008, p. 158) proposes that RST relations are restricted to hold between segments *adjacent to each other* in any direction.<sup>4</sup> The adjacency of segments is defined using the area model (see Section 3.2.2.2). According to Bateman (2008, p. 158), acknowledging the

---

<sup>4</sup>In GeM RST, the base units that contribute to the rhetorical structure are referred to as *segments*.



spatiality as the principle of organisation does justice to its role in the multimodal artefacts. The spatial principle has also found support in eye-tracking studies (cf. Holsanova and Nord 2010, pp. 95-96). Yet RST relations may also extend across layout areas in certain contexts, resulting in what Thomas (2009b, pp. 315-316) has termed “long-distance” relationships in the study of packaging design. It is possible that the tourist brochures will also include long-distance relationships, as their rhetorical structure appears to be increasing in complexity in certain cases (cf. Hiippala 2013). I will thus restrict the relations to adjacent segments, but if the analysis reveals long-distance relationships, they will be accounted for accordingly.

The second challenge relates to the application of RST schema between verbal and visual segments. In total, RST provides five schema for describing relations between segments (Mann and Thompson 1988, p. 247). The two major schema types to be discussed here are nucleus-satellite and nucleus-nucleus. The nucleus-satellite schema is referred to as asymmetric (mononuclear) and the nucleus-nucleus as symmetric (multinuclear). The problem of nuclearity — deciding on schema and whether the verbal or visual element is the nucleus — was previously identified, because the image-text relations appear to be context-dependent and resistant to abstractions (see also Martinec 2013).

Although image-text relations have been extensively explored in both semiotics (Barthes 1977, 1981) and multimodal research (Royce 1998; Martinec and Salway 2005; Kong 2006; O’Halloran 2008b; Liu and O’Halloran 2009), sufficiently reliable and empirically tested frameworks have not yet emerged. What need to be avoided, then, are arbitrary assignments of nuclearity in the image-text relations. Thus, the multinuclear relationship of RESTATEMENT is preferred over forced decisions on nuclearity (Bateman 2008, p. 159).

Finally, the third challenge is related to the image-text combinations on a page. Bateman (2008, pp. 160-161) identifies several cases where the classical RST is not capable of describing the observed phenomena, because the approach lacks the necessary analytical reach. Bateman (2008, pp. 161, 167) supports his argument with several examples in which image and text fragments (from an RST perspective) are connected by lines, but RST does not recognise these fragments as proper rhetorical segments. Instead of representing the relation verbally, these image-text combinations use the two-dimensional space to express the relation. According to the previously introduced principle of spatiality, these relations need to be accounted for in the analysis. For this kind of *subnuclear elaboration*, GeM RST includes five additional relations based on Halliday and Matthiessen (2004) to extend its analytical reach: identification, class-ascription, property-ascription, possession and location (Bateman 2008, p. 162).

This concludes the discussion of the rhetorical layer. In the following section, I shall proceed to describe the final layer to be discussed, that is, the navigation layer.

### 3.2.4 Navigation layer

This section focuses on the navigation layer, which describes the segments that help the reader to interact with the multimodal artefact. These segments may include explicit references, such as ‘see page 5’, page numbers, indices, website addresses, or any other types of reference within or outside of the artefact. The navigation layer is outlined by Henschel (2003), who describes navigation structures in terms of pointers and entries. The pointers, which typically consist of an RST segment, point to another RST segment (entry) or a group of layout units known as a layout chunk (index). The index refers to an entry that identifies a layout chunk, which can be exemplified by the relationship between a page number and the content on the page (cf. Figure 3.2).

Figure 3.5 shows an example of a navigation structure from *Helsinki Your Way* (2006), which uses numbered identifiers that function simultaneously as both pointers and entries. Using numbered identifiers, the navigation structure connects verbal and visual content on different pages: an aerial photograph, a map, a verbal description and a close-up photograph. This multidirectional configuration suggests that navigation structures may take on more complex forms rather than functioning only as pointers and entries. Therefore, the dissertation will use the navigation layer to uncover the use of navigation structures in the tourist brochures, while paying particular attention to their realisation, because in certain configurations the navigation structures seem to be realised implicitly using colour (Hiippala 2012b, p. 1505).



Figure 3.5: Navigation structure in a tourist brochure  
(Source: Hiippala 2012a, p. 119)

Furthermore, the navigation structure of a multimodal artefact may be subtle and appear as naturalised to both users and analysts. In an analysis of a newspaper using the GeM model, Yang notes that “being a printed text, the *Sun Herald* lacks any navigation structure” (2008, p. 306). This misunderstanding may result from the fact that the most prominent form of navigation today takes place in digital media while browsing the internet (see e.g. Djonov 2007, 2008). This kind of *activity*, however, is very different from the navigation *structures* in multimodal artefacts.

In contrast, print media artefacts may contain a wealth of navigation structures to aid the reading process. At the same time, their capability to go unnoticed suggests that the navigation structures need to be carefully analysed. Upon the identification of a certain type of navigation structure, the corpus of the tourist brochures may be searched for other possible matches to expand our knowledge of how the navigation structures are construed in print media.

This section concludes the descriptions of the analytical layers of the GeM model. In the discussion, I emphasised that no layer possesses the descriptive power to independently provide a view of the structure of the artefact. In contrast, the multiple analytical layers need to be brought together for the cross-layer analyses. Moreover, it is equally important to understand the principles behind the phenomena described using the GeM model. For this purpose, I now turn towards a theoretical concept that is geared for this task, that is, the concept of a semiotic mode.

### 3.3 The concept of a semiotic mode

As the previous sections have shown, the GeM model provides information on the structure, organisation and appearance of a multimodal artefact. In this way, the analytical layers allow the identification of structures that may be of interest for a model of a multimodal artefact; for instance, collocations in the layout and the rhetorical structure (cf. Hiippala 2013). However, what is also required is a theoretical tool that can account for these structures and explain how they convey certain kinds of meanings.

For this purpose, I use the notion of a semiotic mode. As Section 2.3.1 established, many theories of multimodality treat language and image as modes which interact in various contexts. This interaction results in the phenomenon of multimodality, which these theories seek to describe. The question is, do these theories and their accounts of mode tell us any more than what the GeM model already reveals about the artefact’s multimodal structure? The answer is undoubtedly yes. Although the GeM model distinguishes between language and image and provides information about their interaction, a more fine-grained definition of mode

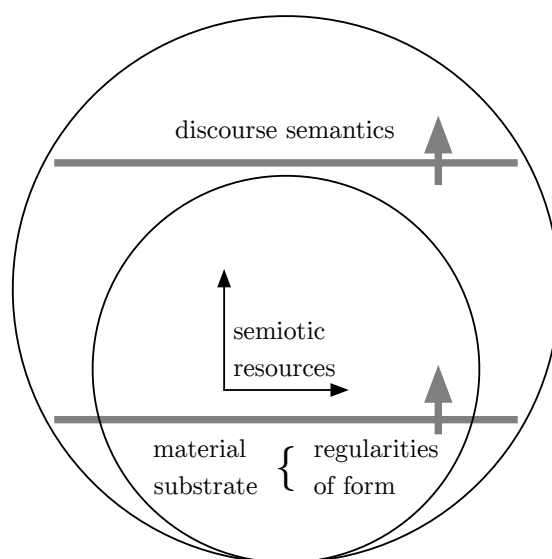


Figure 3.6: The three strata of a semiotic mode  
(Reproduced from Bateman 2011, p. 30)

with greater explanatory power is needed to understand how the tourist brochures communicate multimodally.

I would argue that the definition of a semiotic mode provided by Bateman (2008, 2009*b*, 2011) is the most viable option for this dissertation, due to its compatibility with the GeM framework (cf. the principles of theory-building in Hasan 2004, p. 16). Figure 3.6 shows Bateman’s conceptualisation of the semiotic mode and its three strata: material substrate, semiotic resources and discourse semantics. I will now proceed to discuss each stratum separately in Sections 3.3.1, 3.3.2 and 3.3.3, before describing the semiotic modes active in the tourist brochures in Section 3.4.

### 3.3.1 Material substrate

For Bateman (2011, pp. 20-21), the first stratum of the semiotic mode is the material substrate. To put it simply, a material substrate is a prerequisite for the emergence of a semiotic mode. Without a sufficiently controllable material substrate which may be manipulated by a group of users, the semiotic practices cannot establish a long-lasting form, which prevents their development into a full-blown semiotic mode. Historical examples of material substrates include parchment, papyrus, clay tables, and going back all the way to the Middle Stone Age (250 – 50-25 ka<sup>5</sup> ago) in African prehistory, stone fragments for engravings, whose contribution

---

<sup>5</sup>Kiloannum: thousands of years.

to the rise of “symbolic material culture” and human cognitive development has been recently discussed in Henshilwood and d’Errico (2011a). In short, nothing *permanent*, socially shared and meaningful — as opposed to transient symbolic systems (such as spoken language) — can emerge and persist without a material substrate.

In this dissertation, the material substrate under analysis is paper — and more specifically — the printed page. Although the combinations of language and image on the printed page have been frequently studied in multimodal research, only recent investigations have begun to pay attention to the role of materiality and its relation to mode (cf. Kress 2009; Elleström 2010a), discourse analysis (O’Halloran and Lim 2009) and the semiotics of texture (Djonov and van Leeuwen 2011a). This dissertation, in turn, focuses on the affordances of the material substrate (cf. Gibson 1979), that is, what can be done with the page in terms of semiosis, rather than elaborating the material qualities of the page (see Section 3.5.1). Therefore, the discussion now moves to the semiotic resources, which are carried by the material substrate of a printed page.

### 3.3.2 Semiotic resources

The second stratum of a semiotic mode consists of the semiotic resources, which are “semiotically-charged organisations of material that can be employed for sign-construction” (Bateman 2011, p. 20). Following de Saussure, Hjelmslev and Halliday, Bateman proposes that the semiotic resources should be modelled along paradigmatic and syntagmatic axes, so that the semiotic resources include “paradigmatic systems of choice together with a syntagmatic organisation for re-expressing paradigmatic choices in structural configurations” (2011, p. 20). In simple terms, the semiotic resources allow making choices and combining them in expressions. And for each choice, there are several options from which to choose. This may be illustrated using a simple linguistic example of the paradigmatic and syntagmatic axes of organisation, which is shown in Figure 3.7.

To draw on a non-linguistic example, Seppänen (2005, pp. 128-129) has explored certain syntagmatic and paradigmatic choices in photography. According to Seppänen, syntagmatic choices are made in, for example, aperture, exposure, focal length and lighting, all of which affect the resulting photograph. He relates the paradigmatic choices to the context of situation, which determines the available syntagmatic choices. For example, studio photography affords syntagmatic choices in the setting and the object of photography, for instance, in the background and the clothes worn by the model. The paradigm of news photography, in turn, is unlikely to afford these syntagmatic choices, but the photographer can make other types of syntagmatic choices related to composition and camera angle (see e.g. Caple 2009b).

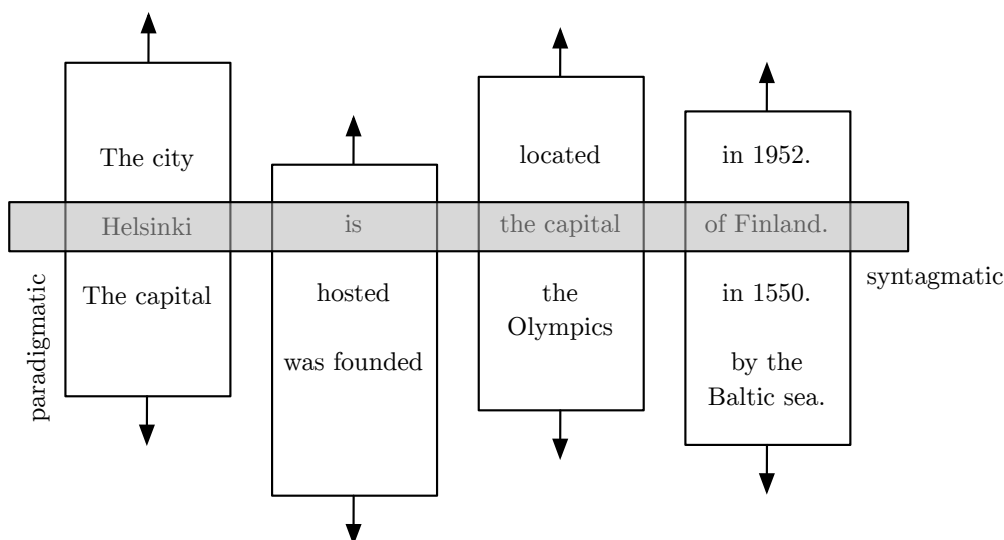


Figure 3.7: Paradigmatic and syntagmatic choices in language (after Bateman and Schmidt 2012, p. 83)

However, a closer inspection reveals that photography encompasses a far more complex paradigmatic and syntagmatic organisation even without invoking the notion of context. Assuming that Seppänen (2005, pp. 128-129) refers colloquially to the camera’s shutter speed as *exposure* (time), the actual exposure is the process whereby light is allowed to fall on a photographic medium. The choices required for an exposure and their syntagmatic and paradigmatic organisation are represented in Figure 3.8. Achieving the desired depth of field requires making paradigmatic choices in the syntagms of aperture, shutter speed and focal length. Moreover, the lighting and the photographic medium also afford a range of choices, which have not been expanded in the system network shown in Figure 3.8.

What is worth noting here is that although Figure 3.8 only describes the physical *realisation* of a photograph — the exposure — the network can be used to realise a photographic representation of any given object in any given situation, and most importantly, in a wide range of different ways. This highlights the immense semiotic potential of photography, which arises from the available choices and their combinations.

However, Bateman (2011, p. 20) points out that each semiotic resource does not necessarily possess a similar structural organisation, drawing on the notions of “grammatically-organised” and “lexically-organised” semiotic resources (Kress and van Leeuwen 2001, p. 113). Whereas grammatically-organised semiotic resources are productive in the sense that they are capable of producing new meanings by combining their sign repertoire, lexically-organised resources have a rather fixed

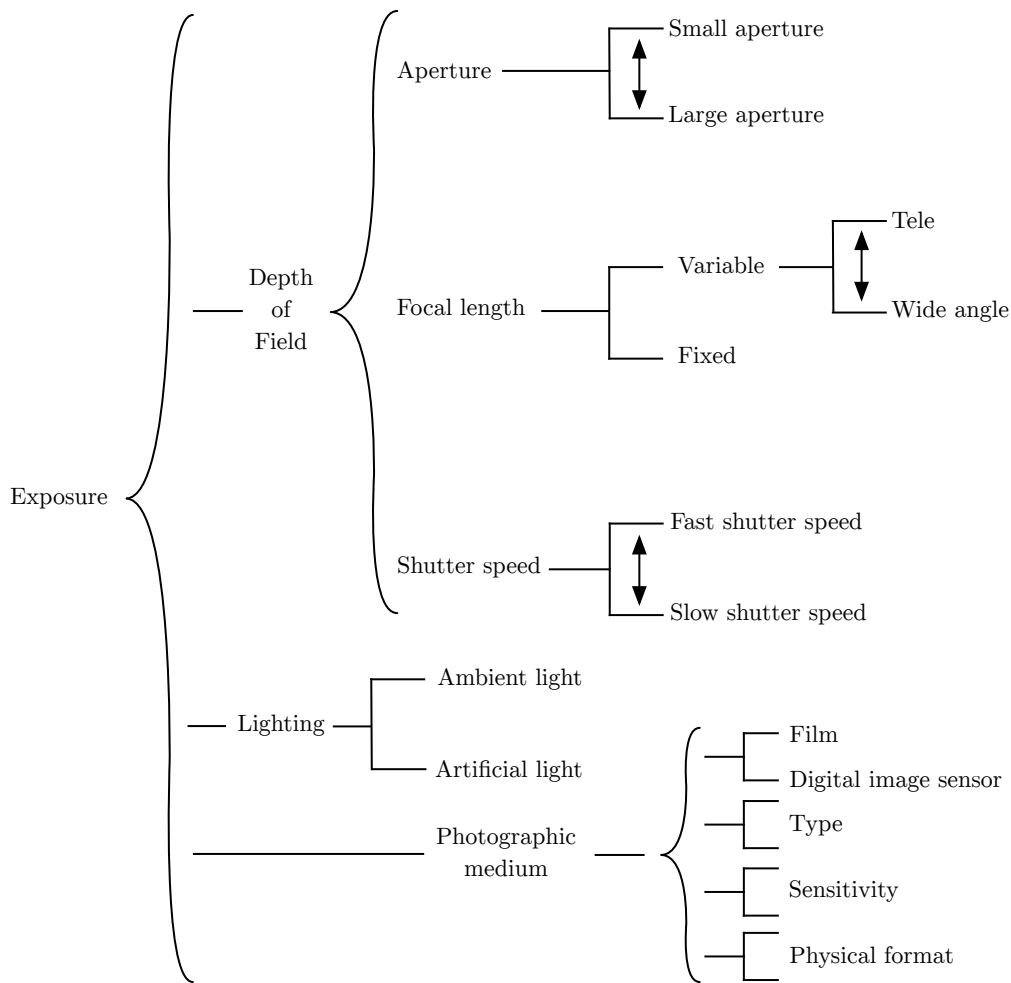


Figure 3.8: Paradigmatic and syntagmatic choices in photography. The two-headed arrows indicate a scale of choices, for instance, in the case of aperture, the  $f$ -number or focal ratio.

sign repertoire. In this connection, it is also important to note that no semiotic resource is “naturally” organised in a particular way: they are shaped by the semiotic needs of a group (Bateman 2011, p. 20).

In addition, individuals may differ in their mastery of a semiotic resource. Consider, for instance, the difference between novice and advanced photographers: the skill of a professional arises from an understanding of the organisation of the semiotic resource and the choices that it allows. Consequently, the advanced photographer’s capability to exploit the organisation of the semiotic resource to make complex meanings exceeds that of the novice.

In short, the creation of aesthetically pleasing, high-quality photographs requires the knowledge of making delicate paradigmatic and syntagmatic choices in the network shown in Figure 3.8. This may be contrasted with automatic exposure, in which the camera makes these choices on behalf of the novice photographer.

Furthermore, I would like to suggest that the gap between the professional and the amateur also defines the photography in the tourist brochures and the photography by the *tourists* (Garrod 2009). As Molina and Esteban (2006, p. 1045) have shown, a visual representation that produces a sense of wonder contributes significantly to the formation of a mental image of the destination. Producing this outcome requires the skill of a professional, who can manipulate the semiotic resource for the desired effect.

Unfortunately, the complexity of the semiotic resources denies their in-depth analysis as a part of this dissertation. For instance, it is unnecessary to emphasise the complexity of language in this connection: the whole field of linguistic science speaks for itself. The previous discussion identified a similar complexity in photographs, which nevertheless represent only one type of visual semiotic resource used in the tourist brochures.

Furthermore, this discussion did not consider the interpretation of photographs in their context of use (see e.g. Machin 2004; Caple 2009*a*). Yet describing the actual use of language and images is necessary for understanding the structure of the tourist brochures. Therefore, I continue by focusing on the semiotic *modes*, which also account for the discourse semantics required for the contextual interpretation of the semiotic resources.

### 3.3.3 Discourse semantics

As stated, the final stratum of a semiotic mode consists of the discourse semantics, which guide the interpretation of the semiotic resources (Bateman 2011, p. 21). It should be noted, however, that this notion is not directly related to the discourse semantics of language, as outlined by Martin (1992), although their names and respective functions are similar. To clarify the role of the discourse semantics, we may draw on the description provided by Bateman (2011, p. 21):

The paradigmatic choices lying behind any particular physical articulation need also to be organised so as to support contextualised interpretation ... that is, the semiotic code<sup>6</sup> only becomes interpretable in context by virtue of being embedded within an unfolding discourse.

---

<sup>6</sup>Bateman (2011, p. 21) defines a semiotic *code* as a combination of a material substrate and the semiotic resources it carries. A full-fledged semiotic *mode*, in turn, requires the discourse semantic component.



Essentially, what Bateman suggests is that the semiotic resources are structured in a way that supports their interpretation in their context of occurrence. This is highly relevant for this dissertation, especially if similar organisations of semiotic resources are encountered in the data. To put it simply, a knowledge of the discourse semantics allows the reader to access the content carried by the semiotic modes, providing an insight into the interpretation of multimodal structures.

Within the proposed model of a semiotic mode, the discourse semantics enable a move beyond viewing language and image simply as semiotic resources. Instead, we may ask *what* the material substrate allows the semiotic resources to do, and *how* the semiotic resources are to be understood as a part of the unfolding discourse during situated discourse interpretations (cf. Bateman 2011, p. 22).

At the moment, the discourse semantics constitute an area that requires further research. This also affects this dissertation: a comprehensive study of their operation in the tourist brochures would warrant another dissertation. For this reason, I will not expand the description of the discourse semantics here. I will return to the topic of representing the discourse semantics shortly in Section 3.5.3 and explore the connection between multimodal structure and discourse semantics later in Chapter 8.

## 3.4 Semiotic modes in the tourist brochures

With the different strata of a semiotic mode now established, we may begin to outline the semiotic modes deployed in the tourist brochures. To set us on the course, Figure 3.9 illustrates three semiotic modes commonly found in print media (see also Bateman 2009*b*). These three semiotic modes — *text-flow*, *image-flow* and *page-flow* — are abstractions which can be used to explain the multimodal phenomena observed in the data. According to Bateman (2008, p. 175), the aforementioned semiotic modes may provide a better understanding of how different semiotic resources are combined in the multimodal artefacts.

Now, to do so, Sections 3.4.1 and 3.4.2 introduce text-flow and image-flow respectively. Page-flow, however, warrants special attention because it may incorporate both text-flow and image-flow. To account for its complexity, the entire Section 3.5 is dedicated to page-flow.

### 3.4.1 *Text-flow*

In its purest form, text-flow is a semiotic mode that consists of one-dimensional, linear-interrupted text (Bateman 2011, p. 26), which may be occasionally interrupted by graphic elements (see Figure 3.9). As a semiotic mode, text-flow is highly adaptive and possesses the entire meaning potential of written lan-

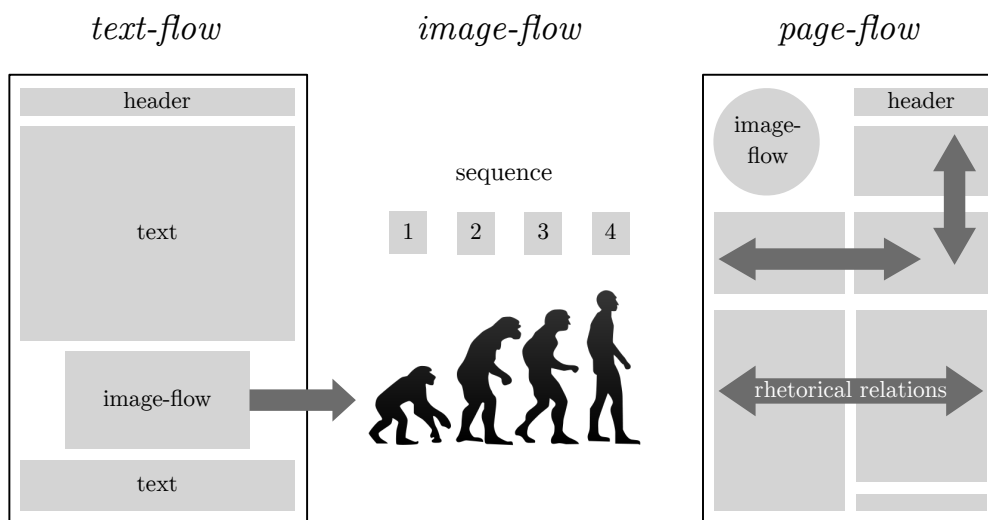


Figure 3.9: Three semiotic modes in printed multimodal artefacts  
 (Source: Hiippala in press)

guage, constrained only by the material substrate and the typographic conventions. The adaptability of text-flow is demonstrated by its wide range of application. Consider, for instance, the difference between this dissertation and the tourist brochures currently under analysis. Although both employ text-flow, the semiotic mode is used for the exchange of different kinds of meanings in a different multimodal environment (cf. Halliday and Martin 1993; Martin and Veel 1998; Hiippala 2007, 2012a).

Unlike page-flow, text-flow does not use the available two-dimensional space to construct additional meanings, but relies on the linear structure of unfolding language (Bateman 2011, p. 26). It is, however, naturally susceptible to a rhetorical organisation and may also contain navigation structures in the form of numbering and references (Bateman 2008, p. 175). Due to its adaptability, text-flow can be expected to be encountered in nearly every instance of data, but the GeM model can capture its rhetorical structure, typographic features, hierarchical organisation, and positioning in a layout.

This is deemed sufficient for this dissertation, as the linguistic features of tourism discourse have been extensively described elsewhere (see e.g. Guijarro and Hernandez 2001; Bonelli and Manca 2002; Kaltenbacher 2006, 2007; Hiippala 2007; Valdeón 2009; Kvåle 2010; Thurlow and Jaworski 2010; Francesconi 2011). This allows me to focus specifically on (1) how text-flow is used independently, (2) which semiotic modes and resources occur within text-flow, and finally, (3) how text-flow operates within page-flow. These issues become focal later in Chapters 7 and 8: the following section continues with a discussion of image-flow.



Figure 3.10: *Bohemian Nordic Oddity* (2006)

### 3.4.2 *Image-flow*

As a material substrate, complemented by sufficiently developed production and printing technologies (cf. Kress and van Leeuwen 2001), the printed page can be used to realise various types of images. To illustrate the use of images in the tourist brochures, we may draw on the previously discussed examples in Figures 2.4 and 3.3. However, working with the notion of a semiotic mode, it has to be kept in mind that the use of images within text-flow and page-flow differs significantly from image-flow, which is a semiotic mode with its own discourse semantics (Bateman 2009*b*, pp. 61-62).

According to Bateman (2008, pp. 175-176), image-flow emerges when images are placed in close proximity with each other, in order to establish a sequential rhetorical relation between them (for an example, see Figure 3.9). Yet all images in the tourist brochures cannot be interpreted this way, which suggests that the images may involve several possible discourse semantic interpretations. Fortunately, as the material substrate of a page only allows the realisation of *static* image-flow (Bateman 2011, p. 26), we can directly exclude its *dynamic* counterpart in film (see e.g. Bateman 2007, 2009*c*; Tseng and Bateman 2010, 2012; Bateman and Schmidt 2012; Wildfeuer 2012).

Nevertheless, our limited knowledge of static image-flow and its relation to other semiotic modes<sup>7</sup> present a significant challenge to this dissertation (cf. Hiippala 2012*c*, pp. 321-323). For this reason, the use of image-flow in the tourist brochures needs to be considered together with the tools currently available for the analysis of image-flow. To exemplify, Figure 3.10 shows a cover of a tourist brochure, which I have treated elsewhere as a narrative sequence depicting nightlife in Helsinki (Hiippala 2007, p. 82), but this interpretation does not arise from the structural configuration of the images. Consequently, the same interpretation is

<sup>7</sup>An example of a semiotic mode that deserves more attention is what could be referred to as the diagrammatic mode, see e.g. Guo (2004).

not necessarily shared by everyone, because the interpretation requires knowledge of the connotative meanings in the images, which arise from the Finnish culture.

Therefore, this kind of detailed, discourse-oriented multimodal analysis remains outside the scope of this dissertation. The decision to treat instances of data as image-flow need to be supported by explicit elements, such as arrows or connecting lines. Given our limited knowledge of image-flow and its communicative functions, it is reasonable to limit the possible interpretations until empirical studies of static image-flow become available. If examples of image-flow are encountered, the participating segments and their sequential relations are annotated using the layout and rhetorical layers.

This concludes the discussion of text-flow and image-flow. Next, I will move to consider page-flow, which may combine both of the aforementioned semiotic modes.

### 3.5 *Page-flow* in the tourist brochures

Bateman (2008, p. 176) introduces the semiotic mode of *page-flow*, which uses the two-dimensional space on the page to organise the content. The main difference between page-flow and text-flow is that page-flow makes use of the entire layout space to work towards the communicative goal of the artefact (see Figure 3.9). For this purpose, page-flow may draw on the output of both semiotic *resources* and *modes*. As Bateman (2008, p. 176) writes:

Page-flow can combine elements in any of the semiotic modes appearing on a page, including text-flow, diagrams, graphs and so on. It adds to the individual contributions of these elements the possibility of a *rhetorical* unity supporting the communicative intentions of the document.

It is imperative for this dissertation to understand how page-flow operates, because the previous research (Hiippala 2007, 2013; Kvåle 2010; Francesconi 2011) suggests that the tourist brochures have moved beyond mere text-flow. On these grounds, we should consider the tourist brochure as a multimodal artefact, whose complexity arises from the potential co-deployment of the three distinct semiotic modes: text-flow, image-flow and page-flow.

The area models in Figure 3.4 may be used to illustrate certain aspects of page-flow. For instance, the cover in Figure 3.4c may be expected to use the layout space in a different manner than the content pages in Figures 3.4a and 3.4b. An index page, in turn, would likely use text-flow, because the navigation structures are realised efficiently using language, whereas the cover may use page-flow with prominent images due to its function as a “contact text” (for a discussion

of covers, see Section 2.5.4). On the whole, these observations require a model of a multimodal artefact to pay specific attention to page-flow and its configuration. For this reason, I will now describe the three strata of page-flow in greater detail.

### 3.5.1 The page as a material substrate

The material substrate of the tourist brochures is the printed page. It does not matter whether the brochure consists of a single page, whether one- or double-sided, or of multiple pages bound together, because each page will have two dimensions, height and width, which form the space in which the semiotic modes may operate. In addition to the physical size of the page, combinations of multiple pages are affected by their *fold geometry* (Bateman et al. 2004, p. 74). When pages are folded and bound together, the area close to the fold becomes less susceptible to visual perception. As Berazhny (2012) has shown, the fold geometry may be exploited to circumvent legal requirements in alcohol advertising by placing the required warnings close to the fold.

Moreover, the interaction with an artefact that combines multiple pages is not limited to the visual perception, but also includes the haptic sense — the sense of touch — which allows the reader to feel and manipulate the artefact. As O’Halloran and Lim (2009, p. 143) have pointed out, the haptic sense may be invoked by altering the material substrate itself, such as the thickness or glossiness of a page, which may also carry connotative meanings (Kress and van Leeuwen 1996, p. 232). Together, the fold geometry and the haptic sense can extend the available layout space, as exemplified by the folding of a magazine centrefold described in O’Halloran and Lim (2009, p. 140). This design solution is also used in the tourist brochures, as I will show in Section 6.1.2.

However, the purpose of the current discussion is not to explore the potential of the haptic sense and materiality, whose semiotic potential is hereby duly acknowledged. The affordances of a page, that is, which semiotic modes may be carried by it, are considered more important for modelling the structure of a multimodal artefact. To continue the discussion, the focus needs to be moved back to the material substrate and how it is utilised. As Section 3.3.1 established, a material substrate is crucial for the emergence of a semiotic mode. In this connection, it is also necessary to consider the tools used to shape the material substrate and to realise the semiotic modes. For this reason, I will now move to consider the production of the tourist brochures.

The GeM model also acknowledges the conditions that affect the production of multimodal artefacts: Bateman (2008, p. 18) defines the fold geometry and physical qualities of the material substrate as *canvas* constraints, whereas the technological aspects of production fall under the category of *production* constraints. The production constraints provide this dissertation with the means to situate the

production of the tourist brochures historically. Obviously, the main impact on the production has resulted from the introduction of desktop publishing (hereafter DTP) (see also Sections 6.3.2 and 8.5).

DTP, which may be broadly described as computer-assisted graphic design, has removed the need to *physically* realise the tourist brochures before production, thus allowing experimentation with the semiotic modes in a digital environment. Moreover, digital printing technology has enabled the realisation of high-quality colour images, which could have been previously constrained by technology and production costs (Bateman 2008, p. 18).

From the perspective of multimodal research, the question that remains to be answered is whether DTP has altered the processes of semiosis, as Manovich (2001, pp. 13-14) has suggested. Therefore, attention needs to be paid to the semiotic modes and their distribution over time. But to do so, we need to learn how to distinguish between the semiotic modes in the tourist brochures. To move forward, I will now continue with a discussion of the semiotic resources and the discourse semantics.

### 3.5.2 The semiotic resources on a page

The page is capable of articulating multiple semiotic resources within the limits set by the canvas and production constraints. As previous studies of page-based artefacts in both linguistic (cf. Martin 1994) and multimodal research (see e.g. Royce 1998; Cheong 2004; Matthiessen 2007; Yang 2008) have shown, a wide range of verbal and visual configurations may be realised on a page. However, Section 3.3.2 established that modelling the entire meaning potential of the semiotic resources is outside the scope of this dissertation (for example, see the partial description of paradigmatic and syntagmatic choices in photography in Figure 3.8). Therefore, I limit the discussion to the framework of semiotic modes and to the elements recognised by the GeM base layer (see Table 3.1).

In the case of language, Section 3.2.1 established that linguistic research has described the structure of tourism discourse to a sufficient degree. Therefore, I describe written language from the perspective of text-flow. And for this reason, the rhetorical organisation and the layout structure constitute the main foci of investigation. However, their contribution to the structure of a tourist brochure is largely unpredictable ahead of the corpus analysis presented in Chapters 7 and 8. Initial explorations in Hiippala (2012*a,b*, 2013) suggest that certain patterns may be observed, but a single analytical layer cannot capture the complex and interrelated choices that shape the semiotic modes.

Moving away from language, the analysis needs to proceed carefully with non-linear linguistic elements (such as tables) and graphic elements recognised as base units in Table 3.1. Photographs, drawings, diagrams, figures and maps exem-

plify the kind of complex linguistic and graphic elements that may occur within text-flow and page-flow. Their complexity is evident in the light of the previous research. Consider, for instance, the extensive work of Barthes (1981), Kress and van Leeuwen (1990, 1996, 2006), Jewitt and Oyama (2001) and Riley (2004) on drawings, illustrations and photographs, O’Toole (1994, 2011) on paintings, MacEachren (1995) on maps, and Kostelnick and Hassett (2003), Twyman (1986), Guo (2004), Baldry and Thibault (2005) and Baldry (2000*a*) on graphs, charts, diagrams and tables. The extent of the previous research shows that expanding the analysis of each graphic or linguistic element is simply not feasible within this dissertation.

It is thus necessary to consider what needs to be known about the semiotic resources when modelling the structure of a multimodal artefact. Holsanova and Nord (2010, p. 83) propose that multimodal artefacts provide the reader with “cue structures” that evoke models of interaction, which enable the recognition of functionally motivated patterns in the artefact.

I would argue that a model of artefact structure needs to uncover these functional patterns and explain them, instead of focusing on deconstructing the individual semiotic resources, which is an equally valuable exercise but with completely different goals. The starting point for identifying these patterns and their functional motivation is the GeM model and particularly its rhetorical and layout layers. However, although the GeM RST allows us to capture the functionally-motivated patterns (see Section 3.2.3.2), a significant effort remains in mapping the discourse semantics of these instances in their context of occurrence (cf. Bateman 2011; Hiippala 2012*c*). The following section continues with a discussion of this issue.

### **3.5.3 Discourse semantics in the tourist brochures**

As Section 3.3.3 established, the semiotic resources become interpretable only in context, that is, within unfolding discourse (Bateman 2011, p. 21). In short, the discourse semantics guide the processes of inference and interpretation and therefore provide the multimodal “cue structures” described by Holsanova and Nord (2010, p. 83). To begin the investigation of page-flow discourse semantics in the tourist brochures, it is necessary to consider the function of the brochures.

Molina and Esteban (2006, pp. 1051-1052) outline the generic functions of the tourist brochures as follows:

A brochure could be defined as a leaflet whose main aim is to motivate people to visit the destination advertised and which includes photographs and the description of the most important monuments and views of the location, as well as information about its services.

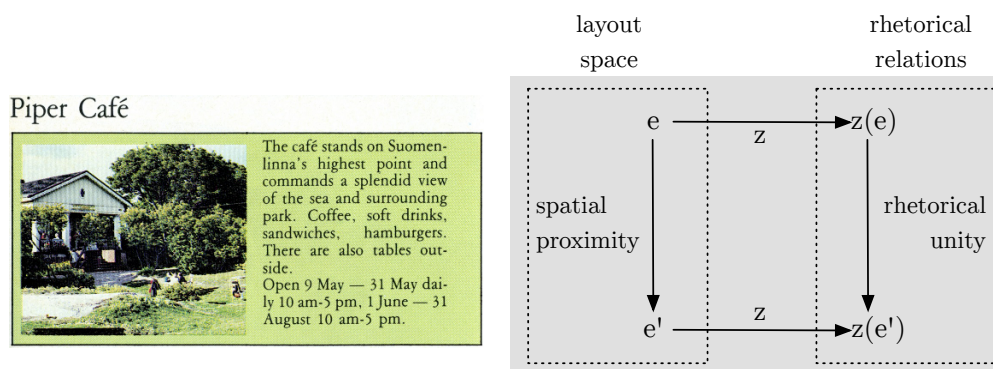


Figure 3.11: An extract from *Suomenlinna Seafortress in Helsinki* (1988)

Previous research (Hiippala 2007, 2012a,b; Kvåle 2010; Francesconi 2011) has shown that the functions described above by Molina and Esteban (2006) may be realised using different multimodal structures. Therefore, I now continue with an example of the discourse semantics, before proceeding to discuss how to distinguish page-flow from other semiotic modes — such as text-flow accompanied by images — in order to gain access to its discourse semantics.

Figure 3.11 shows an extract from *Suomenlinna Seafortress in Helsinki* (1988), which describes a location within the fortress island, along with a representation of its discourse semantics. Firstly, the example itself needs to be considered. In linguistics, the extract's brief verbal description might be described as a “colony text”, because its meaning is not arise from participating in a larger linear linguistic structure, but it is perceived as an independent entity (Hoey 2001, p. 75). Note how the colony text as a non-sequential entity ties in with page-flow, which provides the discourse semantics for interpreting non-sequential organisations of multimodal meaning. Following Kvåle (2010), this combination of verbal and visual elements could be described as an “image-text-complex”. Thus, what needs to be made explicit is how the example promotes the specific interpretation that the verbal and visual elements belong together.

The right-hand side of Figure 3.11 shows a “back-and-forth” diagram, which outlines the discourse semantics for the example on the left-hand side (cf. Bateman 2011, pp. 28-29). The diagram shows two different domains, which are (1) layout space and (2) rhetorical relations. To begin with, the two entities  $e$  and  $e'$  exist within domain (1). Let us assume that in the current example,  $e$  is the photograph, while  $e'$  is the accompanying instance of text-flow. As the diagram shows, a relationship of spatial proximity holds between the two entities, that is, they are placed close to each other.

There also exists a mapping relation  $z$ , which connects the entities  $e$  and  $e'$  in domain (1) with their counterparts  $z(e)$  and  $z(e')$  in domain (2). In domain



(2), which handles the rhetorical relations, the photograph  $z(e)$  and text-flow  $z(e')$  are both considered units with a rhetorical function. As the diagram indicates, a rhetorical relation holds between the entities  $z(e)$  and  $z(e')$ : they are united by a rhetorical structure.

What makes all of the above meaningful is that the mapping relation  $z$  joins the two domains. This allows the spatial proximity of the elements to be translated into rhetorical unity.

The discourse semantics provide a powerful tool for making explicit how the verbal-visual combinations are to be interpreted within page-flow, but their limited reach needs to be acknowledged. In Figure 3.11, the spatial proximity in layout translates to integration in the rhetorical structure. Yet the image-text-complex represents only one possible configuration of multimodal structure, which means that the discourse semantic interpretation presented in Figure 3.11 does *not* hold for all instances of page-flow or for the semiotic mode in general. Therefore, an understanding of page-flow in the tourist brochures requires a careful study of the brochures' multimodal structure. As I will show in the subsequent chapters, the layout and rhetorical layers of the GeM model provide the means to distinguish between the different semiotic modes and their structural configuration.

This concludes the discussion of the three strata of page-flow: the page as a material substrate, the semiotic resources that a page may carry and finally, the discourse semantics. In short, a major challenge for this dissertation is to map the discourse semantics that are required for interpreting the tourist brochures, while connecting them to the patterns observed in the semiotic modes.

Additionally, a model of artefact structure has to be able to formulate certain constraints that set limits to what counts as a tourist brochure (cf. Bateman forthcoming). We may then begin to outline the factors that enable us to recognise the multimodal artefact as a tourist brochure and to invoke the discourse semantic models that are required for its interpretation. However, for the purpose of making generalisations about the multimodal artefacts and the dynamics that affect them, the notion of genre is required. The following chapter, which deals exclusively with genre, provides the final part of the theoretical framework of this dissertation. However, I shall first provide some concluding remarks to the chapter at hand.

## 3.6 Concluding remarks

The aim of this chapter was to discuss the applicability of the *Genre and Multimodality* model to the task undertaken in this dissertation: modelling the structure of a multimodal artefact. By progressing through the analytical layers of the GeM model, I concluded that the model is suited for describing various types of multi-

modal structure. Given that the data of this dissertation is heterogeneous in terms of structure, manifested in different page types and their multimodal configurations, the GeM model provides an appropriate point of departure for studying the data.

I also pointed out that none of the analytical layers of the GeM model — the base, layout, rhetorical and navigation layers — are capable of describing the multimodal structure of an artefact alone. Instead, the so-called *cross-layer analyses* are required to pinpoint the relevant semiotic choices that characterise the tourist brochures as an artefact.

Last but not least, I dedicated a significant part of the discussion to the semiotic modes. The stratified model of a semiotic mode — with its strata of a material substrate, semiotic resources and discourse semantics — is a powerful tool for explaining how the multimodal phenomena described using the GeM model operate. It is important to keep in mind, however, that these semiotic modes are abstractions. If we want to understand how the semiotic modes work, a delicate description using the data provided by the GeM model is required, as Chapter 7 will show.

To conclude, we now possess a framework that enables us to deconstruct and compare multimodal artefacts, and to store this information into a multimodal corpus. The next step is to consider what can be done with the data, and how the data should be interpreted? In the following chapter, I address the notion of genre and how it may help us to understand the data.

# Chapter 4

## Genre

In the previous chapters, I examined and outlined the requirements for a theoretical framework capable of modelling the structure of a multimodal artefact. I concluded that a structure-driven approach to multimodal analysis is the most effective choice for the task at hand. The chosen framework — the Genre and Multimodality (GeM) model — fulfils a key requirement for comparing data, that is, the capability to segment the studied artefacts into analytical units. By building on these analytical units, it becomes possible to compare the structure of multimodal artefacts, provided that the framework is applied consistently to the data.

The questions that follow the collection, analysis and compilation of the data into a corpus are naturally related to data analysis:

- What should we be looking for in the data?
- If we look for patterns, how can they be found?
- If patterns are found in the data, which theories can be used to explain them?

These are the questions which I seek to answer by deploying the notion of *genre* to support the analysis of the data. At this point, however, it should be noted that the following discussion will not touch upon matters such as the representativeness of the data, its analysis and visualisation: they will be addressed in Chapter 5.

The chapter begins with a discussion of genre as a theoretical concept and the challenges of its application in Section 4.1. The remaining chapter has a dual focus, which may be divided along the lines of theory and practice. Sections 4.2 and 4.3 focus on rhetorical, linguistic and multimodal studies of genre, in order to outline the social and communicative functions of the tourist brochures. Section 4.4, in turn, approaches genre from the perspective of document theory and information design, and provides a practical counterpart to the preceding theoretical discussion. Finally, Section 4.5 brings the aforementioned perspectives together into a model of a multimodal artefact.

## 4.1 An outlook on deploying genre

The following sections highlight two important issues in deploying genre as a theoretical concept. Section 4.1.1 explores the challenges that arise from genre as an interdisciplinary theoretical concept. Afterwards, Section 4.1.2 follows up with a discussion of the role of criteria in a definition of genre.

### 4.1.1 On the notion of genre

Any investigation involving genre needs to acknowledge the interdisciplinary nature of the theoretical concept, because different disciplines assign genre with definitions that reflect the discipline's research interests. Consequently, these definitions may be incompatible with each other. The resulting diversity of definitions may be illustrated in the range of disciplines, which have deployed the notion of genre.

These disciplines include, for instance, media and communication studies (Fairclough 1995, 2003; Frow 2006), the study of rhetoric (Miller 1984; Bazerman 1988; Berkenkotter and Huckin 1995), literature (Fowler 1982), library and information sciences (Andersen 2009) and film (Altman 1999). Within linguistics, genre has been deployed as a part of a range of different approaches, such as English for Specific Purposes (hereafter ESP, see e.g. Swales 1990; Bhatia 1993) and systemic-functional linguistics (hereafter SFL, see e.g. Ventola 1987; Paltridge 1997; Christie and Martin 1997; Martin and Rose 2008; Rose 2011; Martin 2012). The emerging field of multimodal research has also applied the concept of genre in the work of van Leeuwen (2005*a,b*), Baldry and Thibault (2005), Lemke (2005), Held (2005) and Bateman (2008, forthcoming), to name a few.

If the discussion is limited to linguistics for the time being, it may be observed that the different approaches also agree on certain characteristics of genre. For example, both ESP and SFL consider genre to be a structured, context-dependent and staged process, which works towards a specific communicative goal. In addition, the organising principle of linguistic genre is considered to be linear. However, with the advance to multimodality, these underlying characteristics need to be reconsidered carefully.

For instance, the principle of linearity becomes problematic — particularly from the perspective of structure — because not all multimodal artefacts are designed to be read in a linear way (Waller 2012). Furthermore, it should be noted that in addition to disciplinary differences, the use of genre may also vary *within* a discipline. Consider, for instance, the following uses of genre in multimodal research:

“Linguistics has seemingly failed to recognise covers as a genre ...”  
(Held 2005, p. 173)

“Advertising is one of the principal genres ...” (Hopearuoho and Ventola 2009, p. 183)

“Newly emergent media such as internet web-pages — an innately hybridic genre ...” (Tan 2010, p. 93)

In the above examples, genre is used to describe three different phenomena: a part of a multimodal artefact (cover), social and economic activity (advertising), and a medium (web-page) — a concept that already carries a considerable theoretical load by itself (Constantinou 2005). Nothing can obviously prevent the use of genre in these contexts, but at the same time, it is also a perfectly valid question to ask what exactly genre contributes to the analysis in these cases (cf. Freadman 2012). As a consequence of the haphazard use, the theorisation of genre within multimodal research has been limited to theoretical discussions (Lemke 2005) or methodological proposals (Bateman 2008).

In some aspects, the situation of genre in multimodal research resembles that of linguistic genre, as described by Bawarshi and Reiff (2010, p. 3):

[T]he term *genre* itself remains fraught with confusion, competing with popular theories of genre as text type and as an artificial system of classification. Part of the confusion has to do with whether genres merely sort and classify the experiences, events, and actions they represent (and are therefore conceived of as labels or containers for meaning), or whether genres reflect, help shape, and even generate what they represent in culturally defined ways (and therefore play a critical role in meaning-making).

This discussion inevitably leads to a point where it is necessary to present the question: what is genre good for in multimodal research?

According to Forceville (2007, p. 1237), genre should complement the detailed multimodal analyses by providing a more abstract level of analysis. He recommends starting with carefully circumscribed corpora, in order to identify the “significant clusters of variables” for each genre. This means that an effective concept of genre should be able to make predictions about the content and structure of a multimodal artefact. Genre should inform the analyst of the artefact’s particular characteristics, their frequency and variation. In this way, the concept of genre would support the detailed analyses, allowing them to fill in the details of the ‘big picture’ of multimodality.

In this chapter, I attempt to formulate a definition of genre that matches the above requirements. Yet to achieve a robust definition, the concept of genre needs to be grounded in the notion of structure. The GeM model, which was described in Chapter 3, is geared towards this purpose and provides the foundation for

this work. The notion of multimodal structure, in turn, needs to be connected to the communicative functions of the tourist brochure. Finally, the multimodal structure has to provide the *criteria* for comparing the tourist brochures: this will be discussed in the following section.

#### 4.1.2 The role of criteria in defining genre

At the core of genre, there needs to be a clearly defined set of criteria for determining its members. As the criteria is used for admission to genre membership, it needs to be rooted in observations made in the artefact structure. Most importantly, the criteria has to be sufficiently constrained. For instance, Martin and Rose (2008, p. 132) discuss different linguistic recount genres, whose key linguistic features they describe as follows:

1. *Personal recount*: serial time; 1<sup>st</sup> person (and 3<sup>rd</sup>); specific participants.
2. *Autobiographical recount*: episodic time; 1<sup>st</sup> person (and 3<sup>rd</sup>); specific participants.
3. *Biographical recount*: episodic time; 3<sup>rd</sup> person (specific); other specific and generic participants.
4. *Historical recount*: episodic time; 3<sup>rd</sup> person; mainly generic participants (but specific ‘great men’).

The above criteria may not only be used to assign a text into a recount genre, but to establish relations between these genres. Consider, for example, the role of time in the recount genres. The first person narrative of a personal recount sets this genre apart from the other genres, in which the narrative episodes take place at different points in episodic time. In short, narrative time provides one possible criterion for establishing the characteristics of a genre.

However, with the move to multimodality, defining criteria for genre membership becomes challenging due to the contribution of multiple semiotic modes. At the same time, the lack of reliable criteria denies us the variables for comparing the artefacts. Therefore, no attempt should be made to identify *a priori* criteria for the tourist brochures. In contrast, the GeM-annotated corpus provides an opportunity to define the criteria from scratch — and based on observation — provided that we can establish the kind of criteria needed. The starting point is nevertheless clear: the analytical layers of the GeM model provide the candidates for the criteria.

Defining the possible sources for criteria may be informed by previous multimodal research. For instance, Kvåle (2010) has studied the role of “image-text-complexes” and their role in describing the activities and locations in the tourist brochures (see also Section 3.5.3), whereas Baldry and Thibault (2005, Chapter 1) have investigated the role of multimodal “clusters” in the formation of genre.

Both of these studies point in the same direction: the content and its organisation in a layout is one possible source for criteria (see also Hiippala 2013). The following step is to consider which layers of the GeM model can be used to investigate the organisation of layout and content. In this case, the investigation should be naturally directed towards the layout and rhetorical layers.

Alternatively, comparing artefacts in the data may involve quantifying a single analytical layer of the GeM model. For instance, Taboada and Habel (2013) use RST to investigate the distribution of rhetorical relations in different multimodal artefacts (for RST, see Section 3.2.3). In the tourist brochures, a similar approach could be used to study the distribution of RST relations in different contexts, by comparing the rhetorical relations to describe destinations, locations and events (Hiippala 2012*a*, pp. 116-117). However, as I will show later, the rhetorical relations are best studied in connection with the other analytical layers for an informed view of discourse structure.

To sum up, carefully-defined criteria plays an important role in successfully applying the notion of genre to the data. Without criteria, the concept of genre is reduced to a mere label. This shall be kept in mind, as I now proceed to discuss previous genre research to establish how it may inform the work undertaken in this dissertation. The discussion begins with the rhetorical studies of genre.

## 4.2 The study of rhetoric: genre as social action

Linguistic genre research has paid special attention to the relationship between language and its contexts of use (Hyland 2002, p. 113). Elsewhere, this relationship has also gained attention in the study of rhetoric. This particular research stream is known by many names: New Rhetoric, North American genre studies and Rhetorical Genre Studies. In the following discussion, I shall use the term Rhetorical Genre Studies (hereafter abbreviated as RGS). Highlighting the differences between linguistic and rhetorical approaches to genre, Hyland (2002, p. 114) describes RGS as strongly rooted in the study of rhetoric and oriented towards ethnographic methodologies, whereas ESP and SFL genre studies mainly work with linguistic text analysis.

The RGS perspective to genre may be exemplified by the work of Miller (1984). In her influential paper, Miller (1984, p. 152) writes:

[I]f genre represents action, it must involve situation and motive, because human action, whether symbolic or otherwise, is interpretable only against a context of situation and through the attributing of motives.

These observations are surprisingly relevant to multimodal research on genre, as there exists a considerable amount of research to support the claim that symbolic exchange — and social action — are inherently multimodal. Whereas spoken language co-occurs with gesture, posture, gaze and so on, written language is combined with typography, images and colour in layouts. Moreover, the interpretation of these multimodal orchestrations is closely connected to the communicative situation or artefact at hand. Miller’s observation underlines the need to develop multimodal discourse semantics, because we also have to understand how the semiotic modes direct us towards particular contextual interpretations (for a discussion of discourse semantics, see Section 3.3.3).

In the light of the above discussion, let us now again consider what kind of communicative work is done by the tourist brochures. According to Molina and Esteban, the tourist brochures constitute “a form of printed promotional material designed to communicate with existing or potential tourists” (2006, p. 1041). They also state that (2006, p. 1050):

[B]rochures should be designed with one of two basic functions in mind: to provide practical information to use in trip decisionmaking and planning processes, and/or to establish an image of the destination as a viable alternative when planning future trips.

The potential reader is likely to have expectations not only towards the content and what the brochures may be used for, but also towards the presentation and organisation of the content (see Section 4.4.1). Together, the two domains constitute a vast field of study, which is unlikely to be adequately covered by a single theory or approach. In this aspect, the potential of RGS lies in large-scale investigations of communicative practices (see e.g. Orlikowski and Yates 1994).

Obviously, the goals of RGS and multimodal analysis differ from each other as well. As Bawarshi and Reiff (2010, p. 59) write:

[RGS] has tended to focus more on how genres enable their users to carry out situated symbolic actions rhetorically and linguistically, and in doing so, to perform social actions and relations, enact social roles, and frame social realities.

Consequently, RGS has considerable potential in ethnographic investigations of multimodality and the use of multimodal artefacts (see Dicks et al. 2011). RGS could be used, for instance, to investigate the role of different artefacts in practising tourism, and to what extent these practices shape the content and form of the tourist brochures. For instance, the tourist brochures encourage the reader to enact the role of a tourist (cf. Culler 1988; Edelheim 2007; Berger 2011) and frame social realities through inclusion and exclusion. For this reason, the content that



disagrees with the communicative goals of a brochure is rarely included in the brochures (Hiippala 2007, p. 24). Multimodal research, in turn, could describe the strategies of presenting the content to the reader.

To conclude, RGS may offer valuable insights on multimodal research, especially if the research is to be extended to a broader context. However, given the goals of the current dissertation, the broad interests of RGS are unlikely to contribute to the practical aspects of modelling the structure of a multimodal artefact. For this reason, I shall now move to discuss the study of genre in ESP and SFL, therefore taking a step towards linguistically-informed analysis.

### 4.3 Linguistic studies of genre in ESP and SFL

This section considers the possible contribution of two linguistic approaches to genre in this dissertation. These two approaches, ESP and SFL, were briefly introduced above. I will now discuss them in greater detail, mainly by drawing on the comprehensive overviews of genre research presented by Hyon (1996), Hyland (2002) and Bawarshi and Reiff (2010).

To begin with, both Hyland (2002, p. 115) and Bawarshi and Reiff (2010, pp. 41-42) draw similarities between ESP and SFL. For both approaches, Hyland (2002, p. 115) emphasises the relationship between the communicative function and formal properties of a text, while Bawarshi and Reiff point out that both share “the fundamental view that linguistic features are connected to social context and function” (2010, pp. 42-43). In addition, both approaches are pedagogically motivated, but the same time, the pedagogical orientation also distinguishes them from one another. Whereas SFL initially focused on the genres used by school-age children (Christie and Martin 1997) and later spread to cover other groups (Martin and Rose 2008), ESP has mainly focused on professional and academic genres (Swales 1990; Bhatia 1993).

As Section 2.2 described, SFL has considerably influenced the development of multimodal theories: I shall therefore mainly focus on the SFL approach to genre. In SFL, genre research has extensively covered spoken and written discourse over a period of several decades. An influential paper in the Firthian linguistic tradition by Mitchell (1957) has inspired SFL genre research and represents an early example of a spoken genre as a staged, goal-oriented process. In the case of Mitchell (1957), the studied genre was a service encounter, which was later described in-depth by Ventola (1987). To sum up, the idea of genre as a series of linear, staged events in unfolding discourse has been central in SFL genre research (see e.g. Halliday and Hasan 1986; Martin 1992, 1994, 1997; Christie and Martin 1997; Martin and Rose 2008; Martin 2012).

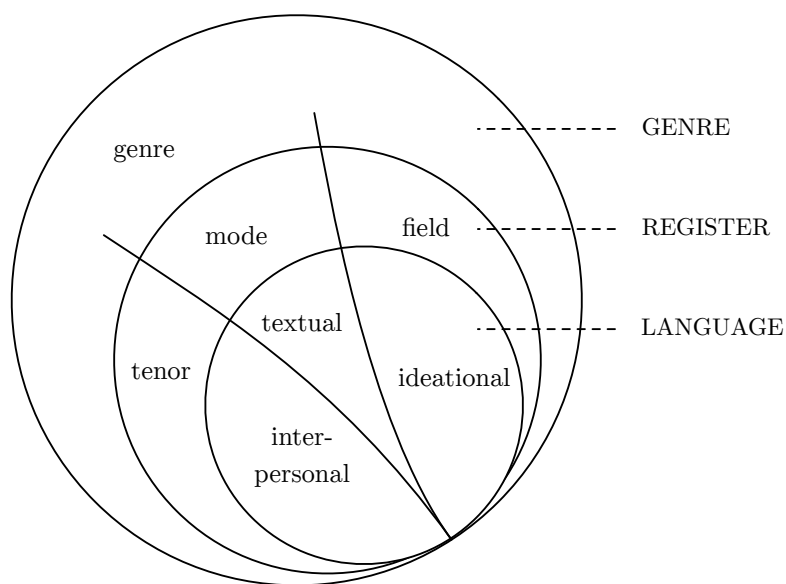


Figure 4.1: Genre in SFL  
(after Martin and Rose 2008, p. 17)

According to one of the prominent scholars of genre in the SFL tradition, James R. Martin, the task of genre is “to account for relations among social processes in more holistic terms” (1997, p. 6). Obviously, these social processes involve the use of language and for this reason, genre is modelled as an abstract semiotic stratum beyond the strata of register and metafunctionally-organised language (see Figure 4.1). The register variables of field, tenor and mode can then be used to generalise about genre on the stratum above, while also accounting for the linguistic choices below (Martin and Rose 2008, pp. 16-17). To exemplify, one may consider the linguistic differences between texts that describe or instruct in the system of MOOD (Martin 1992, pp. 31-32). In this case, the use of indicative and imperative mood reflects the general communicative function of language, which may be then described using the register variables and ultimately using the notion of genre.

In this view, the tourist brochures could employ similar genres, which can be descriptive, instructive or in-between, depending on their intended function (cf. Molina and Esteban 2006, p. 1050). On the other hand, the tourist brochures may also deploy many other genres, such as narratives and news stories, recounts in the form of historical accounts, expositions and discussions (cf. Martin and Rose 2008). The number of potential linguistic genres in the tourist brochures would warrant an independent study: I thus elect not to pursue this line of investigation

here.<sup>1</sup> Instead, I shall turn towards a more relevant issue, that is, how the structure of genre is modelled in SFL.

### 4.3.1 Staging the tourist brochures

As it was pointed out above, SFL considers genre a staged, goal-oriented process with a linear structure. In addition, I briefly mentioned the challenges concerning the principle of linearity in multimodal contexts; a discussion which I shall now continue (see Section 4.1.1). To begin with, the principle of linearity can be used to describe multimodal artefacts, as van Leeuwen (2005*b*) has shown. This, however, requires a considerable loosening of analytical constraints such as replacing the linear description of structure with a linear description of human attention (Bateman forthcoming, p. 4). Nevertheless, certain parts of a multimodal artefact may follow a linear organisation, but the principle should not be overextended due to the reasons presented below. We can, however, take the notion of staging and use it to outline the communicative goals of the tourist brochure.

To illustrate the staging of multimodal artefacts, let us draw on an example from a completely different domain. In a discussion of multimodal genre, Bateman (forthcoming, pp. 11-12) takes up the example of a bird field guide. He argues that the bird field guide entries are subject to generic staging: each entry typically provides background information on the bird, complemented by a description of its physical features and nesting habits. Following Bateman (forthcoming, p. 11), the generic staging of a tourist brochure may also be outlined on the basis of the previous research (Molina and Esteban 2006; Hiippala 2007, 2012*a*; Valdeón 2009; Kväle 2010; Francesconi 2011), as shown in Figure 4.2. In this way, three domains may be identified in the tourist brochures: destination, location and event. Each domain is staged similarly: I will now present the details and the relations that hold between them.

The destination domain covers the entire destination, in this case, the city of Helsinki. The destination is first identified and then followed by a description. The description is typically complemented by information on how to reach the destination and how to obtain answers to further questions. The location domain is similarly staged, but the described locations are embedded within the domain of destination (see Figure 4.2). These locations include landmarks, important sights and services provided. Finally, it is also common for the tourist brochures to describe events that take place within the destination and its locations.

What we also know from multimodal research is that these domains and stages are often realised using multiple semiotic resources: research supporting this view

---

<sup>1</sup>It should be noted that the possibility of studying linguistic genre is accommodated by the GeM model, see Bateman (2008, p. 19).

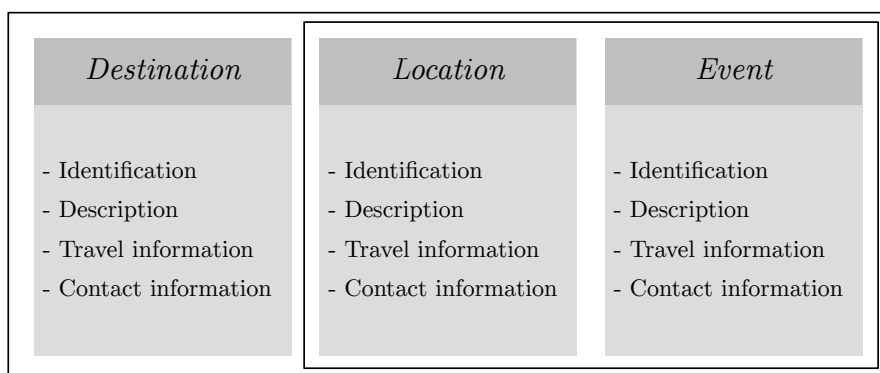


Figure 4.2: A schematic representation of content in the tourist brochures (after Bateman forthcoming, p. 11)

has been reported in several studies (see e.g. Hiippala 2007, 2012*b,a*; Yui Ling Ip 2008; Kvåle 2010; Francesconi 2011). For this reason, applying a model of genre based on linearity needs to be considered carefully, because all artefacts are not designed with the principle of linearity in mind (Waller 2012, p. 239). The readers are aware of this and possess different reading strategies, which are determined by the situation and the artefact in question. In the case of the tourist brochures, it would be a mistake to enforce a strictly linear structure on an artefact that encourages skimming and flicking through the content.

The question is, then, how to capture the ways of presenting and organising the content in the tourist brochures? As opposed to linear models of genre, the semiotic modes may offer a more accurate way of describing how the tourist brochures are structured.<sup>2</sup> Although linearity may be the dominant principle within the semiotic mode of text-flow, the organisation of the entire artefact does not necessarily follow this principle, especially if the artefact uses the semiotic mode of page-flow.

To conclude, an approach that accommodates both linear and non-linear structures is necessary for an accurate picture of multimodality in the tourist brochures. As Chapters 7 and 8 will show, the GeM model and the semiotic modes provide a sufficiently detailed view of the multimodal structures in the tourist brochures. This allows the dissertation to now move on to discuss another relevant issue, that is, the processes of change and their impact on genre and artefact structure.

### 4.3.2 Genre dynamics

Change is an inherent feature of all semiotic systems. The most support for this claim may undoubtedly be found within the study of language, which has described

<sup>2</sup>For a discussion of the key semiotic modes — text-flow, image-flow and page-flow — and their theoretical foundation, see Section 3.3.

linguistic change to a great extent. We also know that linguistic change takes place on several strata ranging from phonetics to discourse semantics. This has to be acknowledged in modelling language evolution, as Rose (2006, p. 87) has pointed out:

Models that do not build discourse systems into their accounts remain unstratified, misconstruing modern adult languages as if they are organized like infant transitional languages, consisting of just two levels, of wordings and their expression as sounds.

Because change also extends to the more abstract strata, such as discourse semantics and genre, it has to be accounted for in longitudinal studies. However, it needs to be conceded that our knowledge of change and its effects on multimodality — and how they are distributed across different semiotic strata — is limited at the time.

What we do know with relative confidence is that change manifests itself as variation. This provides a starting point for investigating change in this dissertation. For this purpose, I shall now move to describe two concepts, *semogenic processes* and *genre agnation*, which may be used to shed light on the changes in the tourist brochures. Finally, I wish to underline that the following discussion considers the application of these concepts *within* the GeM-annotated corpus. Therefore, the concepts are used to describe the relations between artefacts instead of genres.

#### 4.3.2.1 Semogenic processes

The first concept, semogenic processes, may be used to describe the principles of change in semiotic systems — and most importantly — how and where change takes place. According to Halliday and Matthiessen (1999, pp. 17-18), at least three time frames are required to formulate the guiding principles of change. These time frames are now briefly set out, before explaining how they may help us to understand the expansion of multimodal meaning-making. An overview of the time frames is given in Table 4.1.

Table 4.1: Framing semiotic change

logogenesis	‘instantiation of the text/process’	unfolding
ontogenesis	‘development of the individual’	growth
phylogenesis	‘expansion of the culture’	evolution

(Source: Martin 1997, p. 9)

The first time frame — logogenesis — represents the relatively short time frame in which language use occurs. In these logogenetic events, language use is influenced by the knowledge of the overall meaning potential of language, which allows the language user to make the appropriate linguistic choices. Logogenesis is therefore mainly concerned with the use of language as a resource for communication. The knowledge of this resource, in turn, arises from ontogenesis, which is the time frame that accounts for the personal development of a language user.

As said, ontogenesis represents the personal development of the individual's language use. Halliday and Matthiessen (1999, p. 17) point out that this kind of individual development is mainly concerned with growth, not evolution. From this perspective, ontogenesis is concerned with the knowledge of a semiotic system to a lesser or a greater degree (cf. the discussion on photography in Section 3.3.2). In short, ontogenesis represents the totality of experience gathered in events within the logogenetic time frame.

Finally, phylogenesis is concerned with the overall evolution of human language in general, which is shaped by ontogenesis of the individual speakers. This time frame represents the broad, phylogenetic development that enables the evolution of language for specific purposes. For example, the development in the phylogenetic time frame has allowed the emergence of scientific discourse (cf. Halliday 1993; Martin and Veel 1998). In light of the beginnings of *Homo sapiens* as a symbolic species, whose roots are still being uncovered (see e.g. Henshilwood and d'Errico 2011b), this is indeed a considerable achievement, and one which has been made possible by the emergence of media to store and sustain the phylogenetic patterns (cf. O'Halloran 2009a).

As I pointed out above, these time frames are also valid for multimodal research and therefore of high interest to this dissertation (Zhao 2010a; Zappavigna 2010). For instance, let us consider graphic design, which is often characterised as a practice-oriented trade. This is not always the case, however, as Kostelnick and Hassett (2003) have convincingly shown. Although the graphic designers have their individual ontogenetic 'style', the ontogenetic potential is often constrained by the communicative goals. The constraints that bear on ontogenesis closely resemble what the notion of genre attempts to capture in multimodal research. Moreover, it should be noted that these constraints have emerged within the phylogenetic time frame. Unfortunately, as Kostelnick and Hassett (2003, p. 231) point out, the phylogenesis of visual communication remains uncharted:

[The history of visual language] still needs to be written, tracing the evolution of visual language, mapping its genealogy, and analyzing the factors that influenced its development — for example, technology, aesthetics, economics and the discovery of new knowledge.

Consequently, it is necessary to ask what drives phylogenesis? Martin suggests that phylogenesis makes the expansion of meaning possible, which “is a key feature of semiotic systems as they adapt to new discursive and non-discursive (physical and biological) environments” (1997, p. 9) . This makes phylogenesis particularly interesting from the perspective of multimodal research, because it has already touched upon issues which could qualify as adaptation to new discursive environments. What I am referring to here is obviously the emergence and rapid expansion of the digital media. At this point, however, it is difficult to predict whether future will see extensive structural differences in the configuration of semiotic modes in print and digital media (see e.g. Crystal 2001; Bateman et al. 2007). But by building on the model of artefact structure developed in this dissertation, we are better prepared to tackle this challenge in the future.

Finally, it needs to be underlined that capturing meaning-making on the phylogenetic time frame is immensely challenging, if not outright impossible. As Halliday and Matthiessen (2004, p. 47) have correctly noted, the best we can currently do is to use corpora for a snapshot of the semiotic modes at a given time within the phylogenetic time frame. The corpus of this dissertation, which spans over four decades, may thus benefit from the explanatory power of the semogenic processes, especially if specific developments may be detected in the structure of the tourist brochures. Therefore, I shall now move to discuss a possible solution for identifying and representing the structural variation in the tourist brochures.

#### 4.3.2.2 Genre agnation

The second concept, genre agnation, has been used in SFL to model relations between genres. In this context, *agnation* is used to mean relatedness, which is typically manifested in the choices made in language. To exemplify the phenomenon of agnation, let us consider a set of agnate secondary school history genres presented by Martin (1997, p. 15). A partial reproduction of this set is given in Table 4.2.

Table 4.2: Agnate genres among secondary school history genres

Auto / biographical recount	Historical recount	Historical account	Factorial and consequential explanation	Exposition / challenge  Discussion
-----------------------------------	-----------------------	-----------------------	---	---

(Source: Martin 1997, p. 15)

The criteria for the autobiographical and historical recount genres were presented earlier in Section 4.1.2, in which I established that the genres share certain

features (episodic time), while simultaneously differing in other terms (grammatical person, participants). For instance, Martin (1997, p. 15) exemplifies the categorical difference between a historical recount and a historical account by drawing on the way time unfolds in these genres. Historical recounts are organised sequentially, whereas causation is the organising principle of historical accounts. However, Martin emphasises that many texts also combine both principles. This example represents agnation on the most abstract level.

However, Martin and Matthiessen (1991, p. 347) have shown that agnation cuts through the stratified structure of language. On the basis of this observation, it becomes possible to take a step towards considering agnation in terms of multimodality. From this perspective, a multimodal artefact is the outcome of a series of choices in multiple semiotic modes. As Stöckl (2004, p. 15) has noted, the outcome of these choices varies according to the multimodal text or genre in question. Subsequently, multimodal artefacts that work towards the same communicative goals are likely to exhibit similar choices in the semiotic modes, which result in agnate artefacts.

Capturing these agnate choices is a high priority for modelling the structure of a multimodal artefact. This task requires a set of criteria for searching the corpus for alternative configurations in the semiotic modes, and as it was set out in Section 4.1.2, the criteria needs to be based on the GeM model. In this way, it becomes possible to group the artefacts according to the criteria, which may reveal the preference for certain choices. What is thus required is a display mechanism for representing these choices.

So far, several display mechanisms have been explored in linguistic and multimodal research. For instance, Martin (1997, p. 14) uses both paradigms and system networks to model genre agnation. Yet he notes that paradigms become saturated as a display mechanism, as additional criteria are defined (1997, p. 13). From a multimodal perspective, the same applies to system networks, as the choices made in semiotic modes may be very delicate (see e.g. the system network for photography in Figure 3.8). However, if the data in the corpus may be quantified, one suitable candidate for a display mechanism is likely to be a *topology*.

The use of topologies as a display mechanism for representing linguistic structure has been previously explored in Martin and Matthiessen (1991) and Lemke (1999). In particular, Lemke (1999, pp. 2-3) provides a clear definition of topology, which illustrates his background in natural sciences:

A topology, in mathematical terms, is a set of criteria for establishing degrees of nearness or proximity among the members of some category. It turns a 'collection' or a set of objects into a space defined by the relations of those objects.



In this case, the “set of objects” consists of the artefacts in the corpus, whose properties may be contrasted with each other. By quantifying these properties and placing them on two or more axes, the artefacts may be situated in the so-called genre space, which I will describe below.

Bateman (2008, p. 224) has discussed the use of topologies to represent rhetorical strategies, that is, the selection and configuration of the semiotic modes for a communicative goal (cf. Bateman forthcoming). The different rhetorical strategies, positioned along two axes, can be used to establish a genre space. But to operationalise this kind of theoretical construct, we need to be able to quantify the choices made within the semiotic modes. Whether this is already feasible is questionable, because our knowledge of the semiotic modes in the tourist brochures is still limited. For this reason, I will present alternative visual means to study and compare the structure of multimodal artefacts in Section 5.5.2.

This section concludes the theoretical discussion of genre. In the above discussion, I emphasised several issues that need to be considered in applying the notion of genre to the data. These issues included the communicative goals of the brochures, their content and the underlying structures, the definition of criteria for describing these structures, and finally, the processes of change that shape these artefacts. With the theoretical discussion now complete, I shall turn to discuss the notion of genre from a more practical perspective provided by document theory and information design.

## 4.4 Genre in document theory and information design

This section discusses genre from the perspective of *document theory*. For the sake of clarity, I will use the term ‘document’ to refer to the theory and the term ‘artefact’ to describe the object of analysis.

In general, document theory is concerned with the description of “artefact(s) with articulated parts designed to be put back together in various ways by its users” (Bateman and Schmidt 2012, p. 48). This implies that the artefacts are structured in a way that enables the readers to access their content, but this involves interpretative work on behalf of the user. In terms of structure, Bateman and Schmidt (2012, p. 52) propose that artefacts may be generally approached from three broadly defined perspectives. These perspectives are (1) the content view, (2) the logical view and (3) the layout view.

Firstly, the content view is concerned with the content of the artefact. As Bateman and Schmidt (2012, p. 52) point out, much can be said about the content and the way it represents the world. This observation also applies to the tourist

brochures, whose content has been previously studied from various perspectives in multimodal research (see e.g. Hiippala 2007; Yui Ling Ip 2008; Francesconi 2011). For this dissertation, the content of the tourist brochures and its staging were outlined in Section 4.3.1. Yet such in-depth analyses of content do little to advance our understanding of multimodal structure. This requires new perspectives, which I aim to provide in Chapters 6, 7 and 8.

The second perspective — the logical view — describes how the content of an artefact is structured. According to Bateman and Schmidt (2012, p. 53), the logical organisation of an artefact is responsible for several tasks:

This logical view in essence covers part-whole relationships, groups content portions into larger structures of related content, and is typically modelled as a tree structure.

It may be argued that most of the interpretative work by the user involves making sense of the logical organisation of the artefact. The artefact supports this kind of interpretative work by signalling its structure using typography, colour, layout and other means: preliminary research suggests that this kind of signalling reduces the need for cognitive processing (Mayer 2005, p. 192). In the GeM model, the logical view of an artefact is modelled using the layout structure (see Section 3.2.2). As I will show in Chapter 8, the representation of the logical view using the layout structure can tell us many useful things about the organisation of content, especially in connection with the rhetorical structure.

The third perspective is the layout view, which is not to be confused with the GeM layout structure. Adopting a particular layout involves “rendering” the content of an artefact for presentation (Bateman and Schmidt 2012, p. 53). In this case, the material substrate that carries the layout is a sheet of paper. As a part of the rendering process, the logical structure of an artefact adopts specific spatial and visual structures. These structures involve choices made in grouping and placing the content, images, use of colour and typography, etc. In this dissertation, these choices are described using the area model and the realisation information in the GeM model. Their internal logic, in turn, is explained by the semiotic modes.

On the basis of these document theoretical perspectives, it may be postulated that each perspective is subject to *patterning*. The previous knowledge of these patterns acts as the basis for any expectations towards both the content and structure of an artefact. I would argue that these patterns are exactly those which the fields of genre research, document theory and information design have attempted to capture. For this reason, these fields may also complement each other. The following sections will cover the relevant work in these approaches and explore their connections to multimodal research.

### 4.4.1 Genre as an expectation-generating device

One possible starting point for a practice-oriented view of genre is to consider the concept in terms of the expectations it creates. Bateman (2009a, p. 14) offers the following definition of genre, which connects genre to the semiotic modes:

Genre can effectively be characterised as an expectation-generating device: particular combinations of semiotic modes and of particular selections of meanings and forms of expression within those modes are associated with particular tasks taken up within a culture.

A similar perspective — although set out using different terms and from the perspective of information design — can be found in Toms and Campbell (1999), who propose that each document can be treated as an instance of a specific document genre. They refer to the content of a document as its *function*, the layout as its *form* and the material substrate as its *interface* (1999, p. 1).

Both approaches have similar aims, but differ in terms of the theoretical background. Whereas Bateman (2009b) draws on the previous work in Bateman (2008) and anticipates the later work in Bateman (2011, forthcoming), the work of Toms and Campbell (1999) is oriented towards one-time experiment design. As opposed to the theory-oriented multimodal research, the emphasis on the practical work in designing experiments and explaining their results also limits the possible contribution of information design to multimodal research (see Waller et al. 2012, p. 4).

On the other hand, Toms and Campbell (1999, p. 7) also emphasise the need to theorise information design:

Clearly evident from participants' responses and behavior was the concept of taxonomic families of documents based on document structure ... [w]e must adapt concepts from categorization/classification research that define these relationships: are they partitive, suggestive of a part/whole relation, or genetic, suggestive of a parent/child relation?

Keeping the need for a mutually beneficial relation between the different fields of study in mind, I shall now proceed to discuss relevant research in information design, while focusing particularly on the studies which have reached out to multimodal research. First, Section 4.4.2 discusses the work of Waller (2012) on layout and graphic literacy. Second, Section 4.4.3 continues with the work of Waller et al. (2012) on describing information design as a pattern language.

### 4.4.2 Layout as a memory tool

The work of Waller (2012) exemplifies what is hopefully an emerging trend, whereby multimodal research and information design take prospective steps toward each

other. Drawing on studies in information design, linguistics and multimodal research, Waller presents an extensive review of the history and development of layout and its role in multimodal artefacts. This work complements the historical perspectives offered in Twyman (1986) and Kostelnick and Hassett (2003).

Waller (2012) argues that layout plays an important role as a *memory tool*, which allows the reader to access the artefact while simultaneously facilitating its use. In terms of Bateman and Schmidt, layout supports the logical organisation of the content, thus reducing “the need for readers to construct and refer to mental representations of content structure” (2012, p. 53). According to Waller (2012, p. 239), this may be achieved using bulleted lists, diagrams, steps, and other types of formatting, but as I will show in Chapter 8, also by exploiting the interface between the layout and rhetorical structure (see also Hiippala 2013). In addition, Waller (2012, p. 242) highlights the more general role of layout:

Layout is the main signifying feature of many familiar document genres: for example, newspapers, magazines, textbooks, user guides, packaging and reference books.

In relation to the above quote, the role of layout in the tourist brochures needs to be considered. In the tourist brochures, the layout appears to encourage a particular reading strategy; most likely *skimming* or *searching* (Waller 2012, p. 239). Skimming is used for an overview of the document structure, while searching involves finding the answer to a specific question in the document. The encouraged reading strategies need to be in line with the overall function of the brochures: to support decision-making or to communicate a mental image of the destination (Molina and Esteban 2006, p. 1050). Waller (2012, p. 239) sums up the point effectively:

No sensible person chooses from a catalogue, sets up a DVD player, selects a hotel from a travel guide, or looks up a word in dictionary by starting on page 1 and reading through until the end.

This observation raises an important issue about the interpretation of the brochures and their discourse semantics. In Section 3.5.3, I suggested that the brochures often employ the image-text-complexes to describe destinations, locations and events. Figure 3.11 showed how the image-text-complexes translate spatial proximity into rhetorical unity using a specific discourse semantic interpretation. This interpretation, however, holds only for the image-text-complex and does not necessarily apply to all combinations of language and image in the artefact. At the same time, the interpretation of a single page may require working with multiple competing discourse semantic interpretations (Bateman 2011, pp. 32-34).

This implies that additional discourse semantic interpretations may be required beyond the individual image-text-complexes or segments of text-flow, which signal the reader to adopt a suitable reading strategy, using the “cue structures” postulated by Holsanova and Nord (2010, p. 83). These cue structures, which trigger a discourse semantic interpretation, are incorporated into the semiotic modes. For this reason, the interpretation of the entire page may be driven by the organisation of the layout into individual units: I will explore this issue in see Section 8.2.1. Additionally, the very limited experiment described in Hiippala (2013, p. 464) and the more extensive work in Reichenberger et al. (1996) and Toms and Campbell (1999) may be used to increase our knowledge in this area.

In any case, the possibility of *stratified* discourse semantics needs to be accounted for in a model of artefact structure. Practically, this means that the modelling of discourse semantics should not only take place at the lowest and most detailed level, that is, within the child nodes of the layout structure. Instead, the discourse semantic signals that arise from the entire artefact — represented by the organisation of the parent nodes — need to be considered as well (see Section 3.2.2). Keeping the above discussion in mind, I now turn to discuss the general characteristics of information design, as presented by Waller et al. (2012).

### 4.4.3 Document design as a pattern language

The final section covers an important issue related to information design — the logical and spatial organisation of the content and its signalling — and the *practice-based* nature of this process. In a recent paper, Waller et al. (2012, p. 4) set out this issue in detail:

Without anything analogous to “grammaticality” to use as a yardstick, information design tends to rely instead on success measures that are harder to test, such as usability. In practice, rigorous testing with users is often impractical — and so practitioners rely more on “knowing what works” from experience.

Waller et al. (2012) continue by pointing out that without a framework for describing information design — in terms of something analogous to grammaticality — it becomes difficult to express what distinguishes efficient and inefficient design, and how to communicate their respective successes and failures. This observation may be seen as connected to the underlying need for theorisation, whose need was pointed out earlier by Toms and Campbell (1999) in Section 4.4.1.

Essentially, Waller et al. (2012, p. 4) suggest that by building on the genre-based approaches, such as the GeM model, graphic design could be described as a “pattern language”:

In this context, pattern refers not to repeating decorative effects (for example, on wallpaper), but to configurations found consistently within recurring design solutions to common problems. They may be patterns of words, visual configurations, or a combination of both.

The term pattern language, which was coined by Christopher Alexander, originated in the study of architecture. In architecture, design problems are encountered on multiple strata, ranging from furniture, rooms and individual homes to neighbourhoods and entire cities: what a pattern language provides are the solutions to these problems. According to Waller et al. (2012, p. 5), a similar approach could be used to provide solutions to common problems in information design.

In a multimodal context, adopting a pattern to resolve a communicative challenge involves choices in the multiple strata of the semiotic modes. By emphasising stratification, I wish to underline the importance of discourse semantics in relation to any structural pattern in a multimodal artefact. In short, efficient information design clearly signals its user how it should be interpreted. This is precisely how well-working patterns are identified by their designers: the users understand their discourse semantics.

In this case, the most relevant observation of Waller et al. (2012) is that a pattern language may spread out over multiple artefacts and genres that do similar communicative work. Moreover, the proposed approach agrees with the perspective adopted in this dissertation to a large extent — even in terminology. Waller et al. (2012, p. 11) write:

Patterns are also distinct from genres because they are assumed to occur at various different levels of analysis, and many occur across multiple genres (that is, in documents which have very different purposes, content, format, context, etc.).

Within a genre, it is also likely that certain patterns are preferred over alternative options. Waller et al. (2012, p. 11) continue by pointing out that:

If there are some patterns that are most used, most familiar, or more constrained, or that are otherwise considered “best” for a particular genre, we might think of those patterns as the prototypical elements of a genre.

These patterns resemble the “significant variables” of genre, whose study Forceville (2007, p. 1237) encouraged using a circumscribed corpus. What this dissertation attempts to do is to develop the means to capture how the semiotic modes set up these patterns in the tourist brochures. In addition, the dissertation aims to describe the range of possible (and preferred) choices on the basis of the corpus. Next, in the final section of the theoretical framework, I will set out a model of

artefact structure that will be used for this purpose in the analysis, before moving on to describe the data and methods of this dissertation in Chapter 5.

## **4.5 A framework for modelling the structure of a multimodal artefact**

### **1. Determine the functions of the artefact and the used medium**

- The artefact's communicative goals and contexts of use;
- the used medium and its affordances;
- the production, consumption and distribution of the artefact.

These are foundational issues in the analysis of artefact structure that pave the way for the description of content. Chapter 6 discusses these issues in preparation for the next step.

### **2. Determine the content and its structure**

- The semiotic modes used to realise the content;
- the logical and rhetorical organisation of the content,
- the content's typographic and graphic realisation.

The next step is to take apart the contribution of different semiotic modes to the artefact. Chapter 7 presents an in-depth discussion of the content and its multimodal structure.

### **3. Determine the discourse semantics of the content**

- The discourse semantics and their signalling;
- the stratification of discourse semantics.

The final step is to consider how the content is to be interpreted. This is the topic of Chapter 8, which discusses the discourse semantics of the semiotic modes deployed in the tourist brochures, thus bringing the analysis back for a wider perspective to the artefact structure.

# Chapter 5

## Data and methods

The fifth chapter of this dissertation presents the studied data and the methods used for its analysis. Section 5.1 begins with data collection, selection and digitalisation, before describing how this lead to the creation of two overlapping sets of data. I then proceed to discuss the methods used to analyse the data, beginning with the application of the GeM model in Section 5.2. In Sections 5.3 and 5.4, I describe how the GeM-annotated corpus was built and verified, before concluding with a presentation of the query languages and visualisations used to explore the corpus in Section 5.5.

### 5.1 Data

This section describes the data studied in this dissertation. Section 5.1.1 begins with the acquisition and digitalisation of the data, before moving on to consider the criteria for selecting the data in Section 5.1.2. The two data sets and their use in this dissertation are then introduced in Section 5.1.3.

#### 5.1.1 Sources and digitalisation

The main source of data were the Helsinki City Archives. The archives store and maintain annual records and publications by the city, which also include a collection of tourist brochures published by the Helsinki City Tourist & Convention Bureau and its predecessors. The collection covers the period between the early 1960s and early 2000s, but for some reason, the Helsinki City Archives have lost the brochures published during the 1990s. Luckily, the brochures collected by the Finnish National Library were able to fill this gap. The newest brochures (published after 2006) were acquired directly from the Helsinki City Tourist & Convention Bureau. The source for each brochure is given in Appendix A.



The brochures could not be removed from the archives, and therefore the data needed to be scanned. The collected brochures were scanned using an Epson 4490 scanner at a resolution of 300 DPI (dots per inch) and saved in uncompressed JPEG format. This resolution was deemed sufficient for the data analysis and for producing a legible reprint if necessary. In hindsight, a resolution of 600 DPI would have been a better option for a possible application of OCR technology (Thomas 2009*b*, pp. 232-234). In the following section, I will consider the criteria for choosing the data for the dissertation.

### 5.1.2 Representativeness of the annotated corpus

Biber (1993, p. 245) proposes “a reduced set of sampling strata” for an accurate but operationally feasible way for populating a linguistic corpus. This process, of course, is related to the inclusion and exclusion of data. In Table 5.1, I have applied Biber’s “situational parameters” to evaluate whether the collected data agrees with the given operational definitions. The left column gives the parameter; the right column describes the data of this dissertation. I shall now expand the most important definitions below.

To begin with definition (3), the brochures are a form of institutional communication, published by the city of Helsinki. This definition affected data selection, because the publications by other institutions, such as private companies, were excluded. In relation to (4a), the addressees are unenumerated, as no groups or individuals are explicitly specified (Biber 1993, p. 246), and the requirements for using the tourist brochures fall within general knowledge (4d).

Demographic variation within the *producers* of the brochures remains unknown (5a), because no authors are typically acknowledged for an institutional text (5b). Finally, the brochures are at least as factual as advertising is legally required to be (6), they have several purposes (7) and generally discuss a limited number of topics (8). These topics were set out for the genre of a tourist brochure in Figure 4.2.

With the criteria provided by Biber’s parameters, I selected a total of 89 tourist brochures for this dissertation. Using the selected brochures, I created two data sets, which I will now describe below.

### 5.1.3 The two data sets

The collected tourist brochures were turned into two overlapping sets of data. The first data set consists of all brochures that matched the definitions given in Section 5.1.2. Hereafter, I will refer to this data set as *the entire data set*, in order to distinguish it from the second data set. The second data set is the annotated corpus, which consists of the pages selected from the entire data set and annotated

Table 5.1: Situational parameters listed as hierarchical sampling strata

1.	<i>Primary channel</i>	Written
2.	<i>Format</i>	Published
3.	<i>Setting</i>	Institutional
4.	<i>Addressee</i>	
	(a) <i>Plurality</i>	Unenumerated
	(b) <i>Presence (place and time)</i>	Present
	(c) <i>Interactiveness</i>	None
	(d) <i>Shared knowledge</i>	General
5.	<i>Addressor</i>	
	(a) <i>Demographic variation</i>	Unknown
	(b) <i>Acknowledgement</i>	Institution
6.	<i>Factuality</i>	Factual-informational
7.	<i>Purposes</i>	Persuade, entertain, inform, instruct, explain ...
8.	<i>Topics</i>	Destination, locations, events

(adapted from Biber 1993, p. 245)

using the GeM model. In the subsequent discussion, I will refer to this data set as *the annotated corpus*. I will now describe both data sets and their use in this dissertation.

### 5.1.3.1 The entire data set

The entire data set consists of 89 tourist brochures published by the city of Helsinki between 1967 and 2008. The entire data set is given in Appendix A, which provides their names, publication dates, number of pages, and several other features, such as the use of navigation structures, advertisements, and the method of binding. Obviously, these features are too broad for a description of the brochures' multimodal structure and the semiotic modes used. For this reason, it is important to understand that the entire data set is intended for observing the characteristics of the tourist brochure as a *medium* in Chapter 6.

There were other reasons for splitting the data as well. In total, the brochures in the entire data set contain 701 double-pages: the manual annotation of this data was not possible due to the time-consuming nature of the process. Although optical character recognition (OCR) technology has been previously used with the GeM model to produce what Thomas refers to as *proto-GeM* (2009b, p. 243), the proto-GeM annotation created from the OCR output still requires extensive post-processing by humans. Nevertheless, the application of OCR technology in

Table 5.2: The annotated corpus

Year	Name	Identifier	Double-pages
1967	Coming Events in Helsinki	CEH	2
1967	Helsinki's Four Tourist Islands	HFT	2
1967	Museums and Exhibitions in Helsinki	MEH	2
1969	Helsinki: Daughter of the Baltic	HDB	2
1972	Helsinki: Daughter of the Baltic	HDB	2
1972	Helsinki's Four Tourist Islands	HFT	2
1972	Sights In and Around Helsinki	SAH	1
1976	Helsinki's Four Tourist Islands	HFT	6
1977	Helsinki Tourist Information	HTI	1
1978	A Walking Tour in Helsinki (Kruununhaka)	WTZ	2
1980	The Greater Helsinki	TGH	1
1982	Sculptures and Monuments in Helsinki	SMH	2
1983	A Walking Tour in Helsinki (Eläintarha-Töölö)	WTH	2
1984	Helsinki's Four Tourist Islands	HFT	6
1984	Holiday in Helsinki	HIH	1
1986	Suomenlinna Seafortress in Helsinki	SSH	2
1987	See Helsinki on Foot	SHF	1
1988	Helsinki's Four Tourist Islands	HFT	6
1995	Welcome to Helsinki	WEH	1
1998	Welcome to Helsinki	WEH	1
1999	Helsinki Your Way	HYW	1
1999	Helsinki: What, Where, When	WWW	1
2000	Helsinki 1550-2000	H45	2
2001	Suomenlinna Maritime Fortress	SMF	1
2002	Summer Helsinki	SHE	1
2003	Winter-Helsinki	WHE	1
2006	Bohemian Nordic Oddity	BNO	2
2006	Weekend Delight in Helsinki	WDH	1
2008	Groovy Nordic Oddity	GNO	2
2008	Helsinki Visitors Guide	HVG	1

multimodal research is definitely an issue in need of further research, but not within the scope of this dissertation. I thus opted for manual annotation and chose a number of double-pages for the annotated corpus.

### 5.1.3.2 The annotated corpus

To begin with, many of the brochures in the entire data set belonged to the same *series*. By series, I refer to brochures that share the same title and generic content. In a pilot study, I analysed an entire series named *Helsinki's Four Tourist Brochures* and traced the development of its multimodal structure over time (Hi-

ippala 2012b).<sup>1</sup> However, similar descriptions of entire series were not feasible due to the high amount of work required by manual annotation.

Therefore, I decided to sample each series in the entire data set. Upon completion, the 24 series in the entire data set were also represented in the annotated corpus. As Table 5.2 shows, the corpus includes a total of 58 double-pages from 30 tourist brochures. I also included several page types — both covers and content pages — in the annotated corpus to reflect the structural variation in the brochures. Next, I will describe the application of the GeM model to the selected data.

## 5.2 Methods: applying the GeM model

At this point, it is useful to restate several principles behind the GeM model. Firstly, the GeM model is not a prescriptive model of multimodal structure. Instead, it is intended to provide a set of analytic tools that may be applied to any instance of data. This set of tools is not considered exhaustive, but sufficient for capturing the basic properties of multimodal artefacts (Bateman 2008, p. 15). Secondly, a successful application of the model results in a creation of a multimodal corpus. But in order for the corpus to be useful, the model needs to be applied consistently. Thirdly, theories that may explain the phenomena encountered in the corpus are required. These are provided by the notions of semiotic mode, genre and other associated concepts described in Chapters 2, 3 and 4.

The following section gives a layer-by-layer description of the GeM model and its application to the data. As described in Chapter 3, the analytical layers of the GeM model account for the different structural aspects of the multimodal artefacts. The annotation for each analytical layer is stored in a separate XML file, but these files are cross-referenced in the annotation using identifiers, as indicated in Figure 5.1.

As Figure 5.1 shows, the base layer is central to the model because all other analytical layers cross-reference the layer. This kind of layered, cross-referenced approach is useful for querying the corpus, because the entire information stored in the corpus becomes available through cross-referencing. Patterns may then be sought not only within, but also across the layers. However, the role of each layer needs to be understood in order to design effective queries. The combination of information from the different layers is what grants the GeM model its analytical potential, because every layer focuses on a different aspect of artefact structure. Therefore, I will now proceed to describe the application of each analytical layer below.

---

<sup>1</sup>The pilot study was conducted at the University of Bremen during spring 2009.

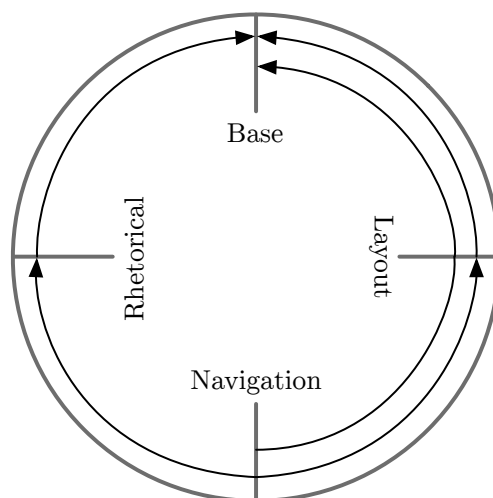


Figure 5.1: The cross-referencing of GeM annotation

### 5.2.1 Data segmentation: base layer

The first step was to segment the data into analytical units. This created the *base layer*, which provides the units for other analytical layers (see Section 3.2.1). Following the list of Recognised Base Units (RBUs) set out in Table 3.1, the base units were annotated and stored in XML files. Each individual base unit was given a unique identifier, for example, *u-1.01*. The prefix ‘u-’ indicates that the identifier belongs to the base layer. In the following digits, ‘1’ indicates the double-page where the base unit is located. The number ‘01’, in turn, is the individual identifier for the base unit. These identifiers were assigned in the order of analysis (see Table 5.3).

Table 5.3: Base layer elements and identifiers

Element	Identifier
<unit>	u-1.01, u-1.02 ... u-2.01, u-2.02 ...

The base units also store the content of the brochures. For instance, the brochure *Helsinki’s Four Tourist Islands* included the following sentence: “Seura-saari also has a restaurant and beaches”. This sentence was stored in *unit* element with the identifier *u-2.14*. In addition, the brochure also included a photograph of a beach. It is clear that a photograph cannot be included in the corpus in the way a sentence is stored between the *unit* elements, yet it needs to be accounted for in the analysis. Therefore, for photographs and other graphic elements, the model uses the attribute *alt* within the *unit* element for an alternative description. The

following labels were used to describe the graphic elements: photo, map, illustration, logo, diagram, and line. The above examples above would thus appear in the corpus as follows:

```
<unit id="u-2.14">Seurasaari also has a restaurant and beaches.</unit>  
<unit id="u-2.33" alt="Photo: Beach"></unit>
```

Because the content is stored in the *unit* elements in the base layer, the other layers of the GeM model do not have to store the same content. Instead, a simple reference to the unit identifier in the *id* attribute suffices to retrieve the content. With the content segmented into analytical units, it became possible to assign them with additional information. At this point, the corpus contained plain verbal text and brief descriptions of the images. However, a model of the multimodal structure requires more information: the location of the base units in the layout, their typographic and graphic features, their relationship with other base units, and so on. This was achieved using the analytical layers, which are described next, beginning with the layout layer.

## 5.2.2 Composition and appearance: layout layer

The primary task of the layout layer is to account for the organisation and appearance of the content. As stated above, the content is stored in the base layer. In practice, the description of *organisation* includes two aspects: (1) tracking the spatial position of elements and (2) defining their hierarchical organisation. Describing the *appearance* of the data, in turn, involves an account of the typographic and graphic features of the content. Together, the different components of the layout layer provide a comprehensive view of the content's organisation and appearance.

The first step in applying the layout layer was to segment the base units into layout units: the principles behind this process were described in Section 3.2.2. Both the layout structure and realisation information were considered in assigning base units into a layout unit, as shown in the annotation example below:

```
<segmentation>  
  <layout-unit id="lay-1.01" xref="u-1.01 u-1.02 u-1.03 u-1.04 u-1.05">  
    <layout-unit id="lay-1.01.1" xref="u-1.01"/>  
  </layout-unit>  
</segmentation>
```

The layout units constitute the basic unit of the layout layer. These units, which are cross-referenced with the base layer, can be then described using the components of the layout layer. These three components — layout structure, realisation information and area model — include the elements required for describing

Table 5.4: Layout layer elements and identifiers

Element	Identifier
<layout-root>	id-year-l-cover, id-year-l-pages-1-2, id-year-l-pages-3-4 ...
<layout-chunk>	freely defined unique identifier, preferably descriptive
<layout-leaf>	layout-unit id
<area-root>	id-year-a-cover, id-year-a-pages-1-2, id-year-a-pages-3-4 or id-year-a-side-1
<sub-area>	page-a-back, page-a-front, page-a-1, page-a-2, page-a-3 ...
<sub-area> (nested)	page-a-1-b, page-a-1-c ... page-a-1-b-1 ...
<text>	layout-unit id
<graphics>	layout-unit id

the various aspects of layout, which I will now describe in greater detail. These elements and their identifiers are given in Table 5.4. I shall begin with a description of the area model.

### 5.2.2.1 Area model

The area models, which were introduced in Section 3.2.2.2, were created using two graphics editing tools. This process involved two separate steps:

1. Reconstructing the scanned brochures and drawing a representation of the layout grid in Corel Paint Shop Pro X<sup>2</sup> (hereafter PSP).
2. Measuring the height and width of the identified areas using the measurement tool in GNU Image Manipulation Program<sup>3</sup> (hereafter GIMP).

In some cases, the scanned brochures were too large for the glass surface of the scanner and had to be scanned and stored in separate files. These files were then joined together in PSP, which was also used for the area model by drawing a layer of lines to represent the baseline grid used to aid the placement of content. Figure 5.2 shows how the lines are used to indicate the areas into which the content was placed. When the underlying image of the brochure was removed, the lines remained and provided a representation of the layout and the used baseline grid. GIMP was then used to measure the physical size of the layout areas in millimetres.

Let us now look at an example of the GeM annotation for the area model shown in Figure 5.2. To begin with, the *area-model* element contains the *area-root* element with various attributes which are now discussed, because they explain the

<sup>2</sup><http://www.corel.com>

<sup>3</sup><http://www.gimp.org>

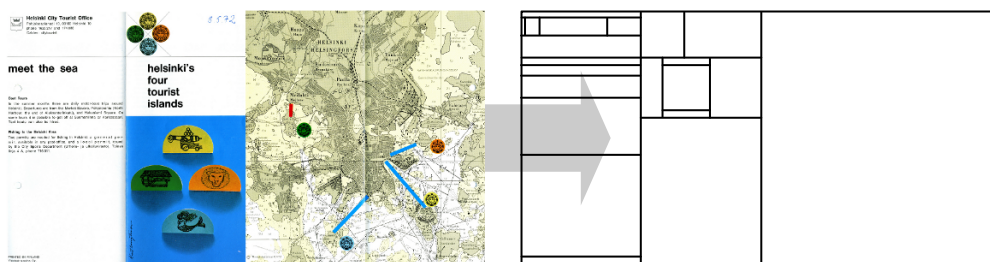


Figure 5.2: Area model of *Helsinki's Four Tourist Islands* (1972)

principles of annotating the area models. Firstly, the unique identifier of the area model in the *id* attribute distinguishes the area model from other instances in the corpus. In this case, the identifier was *hft-1972-a-side-1*.

```
<area-model>
<area-root id="hft-1972-a-side-1" cols="3" rows="1" hspacing="25 25 50"
  vspacing="100" width="404mm" height="210mm">
  <sub-area id="page-a-1" location="col-1" cols="1" rows="9"
    hspacing="100" vspacing="2 8 9 3 3 10 22 39 4" width="101mm"
    height="210mm">
    ...
  </sub-area>
</area-root>
</area-model>
```

The area model also includes attributes that describe the physical characteristics of the analysed brochure. The attributes *width* and *height* contain the information collected using the measurement tool in GIMP. For instance, the example in Figure 5.2 is 404 millimetres in width and 210 millimetres in height, which is annotated as exemplified above, while the attributes *vspacing* and *hspacing* define the grid. The content in Figure 5.2 was organised into three columns, which can be identified by the headers “meet the sea” and “helsinki’s four tourist islands” and the map.

The horizontal division of the grid is described using the *hspacing* attribute, which gives the percentages of the area model: 25, 25 and 50. Considering that the width of the brochure is 404 millimetres, we can then calculate the width of each column using the values given in the *hspacing* attribute. It should be noted, however, that in this case, there is no vertical grid structure at the *area-root* level. Thus, the *vspacing* attribute has the value of 100 percent. This provides a rough outline of the area model with three columns and a single row that is not, however, sufficient for establishing the exact location of each content element in the brochure.



The next step is to define the *sub-area* elements that elaborate the initial layout defined in the *area-root* element. As the example above shows, the *sub-area* element shares many of the same attributes with its parent element *area-root*. Additional information is provided by the *location* attribute, which states that the sub-area is located in the first column of the layout area defined in the *area-root* element (see Figure 5.2). Furthermore, the *rows* and *vspacing* attributes indicate that the sub-area contains a total of nine rows, whose size may be calculated using the *height* attribute. Finally, additional nested *sub-area* elements may be defined to further elaborate the grid.

To conclude, the successful application of the area model requires that each identified layout unit can be assigned with accurate location information. In addition to providing reference coordinates using the grid, additional location information could be defined in the layout structure annotation, which I will describe shortly below. However, certain issues about annotating the layout layer need to be raised before proceeding.

Firstly, the measurements were not always precise, because they were not measured from the original prints on a flat surface. The scanning of bound artefacts may have resulted in distortions on a scale of several millimetres, depending on how the brochure is bound. Secondly, the GeM schema currently lacks the means to annotate elements with non-rectangular shapes, as their calculation by hand would require considerable effort and an expansion of the annotation schema (for an example, see Figure 3.4c). Therefore, the positions of non-rectangular elements were also defined using a rectangular grid.

With the description of the area model complete, I will now proceed to discuss the layout structure.

### 5.2.2.2 Layout structure

The layout structure serves two distinct purposes. Firstly, it describes the hierarchical organisation of the content, and secondly, it also accounts for the placement of content in the area model. As the example below shows, the layout structure references the area model in the *area-ref* and *location* attributes, thus establishing a relationship between the two layers (cf. Bateman 2008, p. 126).

```
<layout-root id="hft-1972-1-side-2" orientation="landscape">
  <layout-chunk id="page-1-side-2" location="multi" area-ref="page-b-1">
    <layout-chunk id="header-page-1-side-2" location="row-1" area-ref="page-b-1">
      <layout-leaf xref="lay-2.01.1" location="row-2" area-ref="page-b-1-b-1"
        valign="left" halign="top"/>
      ...
    </layout-chunk>
  </layout-chunk>
</layout-root>
```

The layout structure also provides additional information on the spatial positioning of the content using the *valign* and *halign* attributes. These attributes are used to

indicate if the content is aligned with a particular side of the grid defined in the area model. For example, a layout unit consisting of verbal text can be aligned with the left and top sides of the layout area. The ‘helsinki’s four tourist islands’ heading in Figure 5.2 exemplifies this case: the text is justified left and begins at the top of the sub-area.

To conclude, the layout structure provided a valuable perspective into the logical organisation of the content. By itself, the contribution of the layout structure may appear as modest. However, a combination of the layout structure and the rhetorical structure provided the means to describe how the parts of the artefact establish their functions within the overall structure, as I will show later in Chapter 7. I will now proceed to describe the final component of the layout layer: the realisation information.

### 5.2.2.3 Realisation information

The final component of the layout layer is the realisation information, which was responsible for describing the appearance of each base unit identified in the analysed artefact. For this purpose, the GeM model includes a schema for describing the typographic and graphic features of the base units using two different elements: *text* and *graphics*. Both elements have their individual attributes that describe their realisation features.

The realisation information is assigned using the layout layer, as the following annotation example shows. The *xref* attribute in both *text* and *graphics* elements is used for identifying the described layout unit. Additionally, the *hi-text* element, which possesses the same attributes as *text*, is used to indicate typographic variation in the layout units. The *hi-text* element is needed, for example, to indicate the use of a bold typeface to highlight a single word or phrase within a paragraph.

```
<text xref="lay-1.01.1" font-family="sans-serif" font-size="10"
  font-style="normal" font-weight="bold" case="mixed" color="black"/>
<graphics xref="lay-1.01.2" type="illustration" colors="black"
  width="48mm" height="28mm"/>
<text xref="lay-1.01.3" font-family="sans-serif" font-size="8"
  font-style="normal" font-weight="bold" case="mixed" color="black"/>
...
```

The typographic features were measured using the measurement tool in GIMP. The unit of measurement was a point (pt), which are given in the *font-size* attribute. For an extensive list of other attributes for describing the *text* element, see Henschel (2003, p. 6). I will not further elaborate on the annotation and attributes of typographic realisation here, because their description is already relatively developed and was thus considered suitable for the analysis of the tourist brochures (cf. Bateman 2008, p. 121).

Table 5.5: Entry types for the *type* attribute in the *graphics* element

illustration	photo	diagram	two-d-element	map
--------------	-------	---------	---------------	-----

At this point, it was necessary to expand the original GeM model to accommodate an additional graphic element. The additional entry type ‘map’ was defined for the *type* attribute in the *graphics* element. This entry type was used to describe maps, which are a common feature of the tourist brochures (see Section 7.3.3). The updated list of allowed entries for the *type* attribute is given in Table 5.5.

The entry type ‘two-d-element’ which covers a range of two-dimensional elements, such as arrows, lines and icons, requires additional attributes to describe their properties (Henschel 2003, p. 7). Finally, the application of the various entry types to the data needs to be made explicit. The following list provides the broad criteria for annotating the graphic elements using a particular entry type:

- **illustration:** hand- or computer-drawn illustrations, logos, etc.
- **photo:** photographic images
- **diagram:** information graphics
- **two-d-element:** lines, arrows, icons, etc.
- **map:** graphic elements communicating geographical or spatial information

The size of these graphic elements was measured using GIMP in millimetres and annotated using the *width* and *height* attributes. This concludes the description of the layout layer: I will now continue with a description of the rhetorical layer.

### 5.2.3 Presenting the content: rhetorical layer

Section 3.2.3 introduced the rhetorical layer of the GeM model, which uses an application of Rhetorical Structure Theory (RST) to describe how multimodal artefacts present and argue for their content. Essentially, the rhetorical layer describes how different parts of the artefact relate to each other using a set of rhetorical relations, which attempt to capture the intentions of the designer(s) in both verbal and visual communication.

In this dissertation, the rhetorical layer was used to describe the multimodal discourse structure of the tourist brochures, that is, how the linguistic and graphic elements work together to form a coherent instance of multimodal discourse. In short, I used the rhetorical layer to deconstruct the discourse structure of the tourist brochures, and to observe any patterns that may emerge, particularly in connection with the previously described layout layer.

The application of the rhetorical layer began with a process of segmentation, in which the units that contribute to the rhetorical structure were selected from the base layer. These units were annotated using the *segment* element and its *xref* attribute:

```
<segment id="s-1.07" xref="u-1.08">Boat tours</segment>
<segment id="s-1.08" xref="u-1.09">In the summer months there are daily
  motor-boat trips around Helsinki.</segment>
```

The *id* attribute of the *segment* element was then picked up in the RST relation definitions, which are given in Appendix B. Three elements were used for annotating the asymmetric, symmetric and subnuclear relations:

```
<span id="span-1.02" nucleus="s-1.01" satellites="span-1.03"
  relation="enablement"/>
<multi-span id="span-1.03" nuclei="s-1.03 s-1.04 s-1.05" relation="joint"/>
<mini-span id="span-1.04" attribuend="s-1.06" attribute="s-1.07"
  relation="identification"/>
```

The *multi-span* element is used for symmetric relations with two nuclei, whereas its asymmetric counterpart uses the *span* element. Both elements, which indicate the relationship holding between the nuclei and satellites using the *relation* attribute, draw on the initial segmentation performed in the rhetorical layer for the analytical units. The *mini-span* element, in turn, is used to indicate the use of subnuclear relations. The elements and their identifiers are given in Table 5.6.

Table 5.6: Rhetorical layer elements and identifiers

Element	Identifier
<rst-structure>	None
<mini-structure>	None
<segment>	s-1.01, 1.02 ... s-2.01, 2.02 ...
<mini-segment>	s-1.01, 1.02 ... s-2.01, 2.02 ...
<span>	span-1.01, span-1.02 ... span-2.01, span-2.02 ...
<multi-span>	span-1.01, span-1.02 ... span-2.01, span-2.02 ...
<mini-span>	span-1.01, span-1.02 ... span-2.01, span-2.02 ...

Finally, I will not present separate ‘use cases’ for the rhetorical relations in this connection, because their use will be described in detail in Chapter 7.

## 5.2.4 Guiding the reader: navigation layer

The final layer to be discussed is the navigation layer, which accounts for the structures that help the reader to use the artefact. It should be noted that these

structures, such as page numbering and indices, were regarded as a feature of the used medium (see Chapter 6). This means that the medium of a brochure has the potential to include a navigation structure, a feature which is shared by many other print media as well (cf. Bateman forthcoming, p. 12). Consequently, a tourist brochure may include a navigation structure as a part of the artefact.

Table 5.7: Navigation layer identifiers

Element	Identifier
<pointer>	p-1.01, p-1.02 ... p-2.01, p-2.02 ...
<entry>	e-1.01, e-1.02 ... e-2.01, e-2.02 ...

As Figure 5.1 shows, the navigation layer includes references to two other layers in the GeM model: the base layer and the rhetorical layer. The annotation for the navigation structure, in turn, involved two elements: *pointers* and *entries*. The identifiers for these elements are given in Table 5.7 and exemplified below.

```
<pointer id="p-2.01" type="cross-reference" from="s-2.02" to="e-1.01"
  range="document-internal">
  <content xref="u-2.02"/>
  <address xref="e-1.01" address-precision="unsure"/>
  ...
<entry id="e-1.01" xref="u-1.21" rst-span="s-1.17"/>
```

The *pointer* element draws on the rhetorical layer for its point of origin in the *from* attribute. The *to* attribute, in turn, is defined in the navigation layer and refers to a defined *entry* element. The *type* and *range* attributes of the *pointer* element define the pointer type and its range (see Henschel 2003, p. 23), that is, whether the pointer points to an entity within or outside of the artefact. Furthermore, the *pointer* element contains two additional elements, *content* and *address*, which cross-reference the base layer and provide additional information on the *entry* element targeted by the *pointer*.

The *entry* element references both base and rhetorical layers. The *xref* attribute identifies the base unit that acts as the entry point, while the attribute *rst-span* identifies the referred rhetorical segment.

This concludes the description of how the GeM model was applied to the data. In the following section, I will continue with a description of the GeM-annotated corpus and how the corpus was built and stored in XML files.

### 5.3 Building the corpus

This section describes how the files containing the GeM-annotated data were compiled into a corpus and how the annotation was verified.

### 5.3.1 File names

The GeM model stores the annotation in multiple files, which means that the annotation for each analytical layer is stored in its own file. As a rule of thumb, the GeM model requires four files for each artefact, but each file may contain the annotation for multiple pages of the same artefact. This results in multiple files and requires that these files can be distinguished from each other.

For each brochure, I used a simple naming procedure, which is exemplified below using the file names for the brochure *Helsinki: Daughter of the Baltic* (1969):

- Base layer: *1969-hft-base-1.xml*
- Layout layer: *1969-hft-lay-1.xml*
- Rhetorical layer: *1969-hft-rst-1.xml*
- Navigation layer: *1969-hft-nav-1.xml*

As the list shows, each file name follows a pattern: year, identifier, layer name and the number 1. The used identifiers are given in Table 5.2. The number 1 is a legacy element from the initial annotation, which stored the annotation for each double-page in a separate file. However, this did not prove feasible, but I retained the identifier for possible future work involving different annotation schema, for example, an application of classical RST or a comparison between manual and automatic annotation. Files with different annotation schema may then be identified using the final number in the file name. Next, I will describe how the annotation was verified.

### 5.3.2 Document Type Definitions (DTD)

The Document Type Definitions (DTD) play an important role in the annotation process: they define the allowed elements and attributes for each layer of the GeM model. For this reason, the DTDs can be used to verify whether the annotation follows the markup declarations given in the DTDs.

The DTDs used in this dissertation were defined using RELAX NG (REgular LAnguage for XML Next Generation) schema language,<sup>4</sup> compiled and first used by Martin Thomas (2009b). We collaborated on making minor changes to the DTDs: for example, we introduced the map as a new *graphics* type to the realisation information.

However, the DTDs cannot prevent human errors in the annotation: I thus implemented additional verification measures, which are now described below.

---

<sup>4</sup><http://www.relaxng.org>

## 5.4 Verifying the corpus

The manual annotation of corpora always carries the risk of human error. In this context, I do not refer to analytical errors in applying the GeM model, but mainly to errors in the identifiers responsible for cross-referencing and identifying the analytical units. The DTDs described in Section 5.3.2 can prevent some of these errors, but not all of them. If the markup is valid XML, the DTDs cannot detect if the identifiers contain errors.

Therefore, the only way to locate these errors was to verify the annotation manually. Due to the amount of data, the verification was done with the help of visualisations, which I will describe in detail in Section 5.5.2.

Thomas (2009*b*, p. 270) was the first to use visualisations to verify the annotation in a GeM corpus. He identified several annotation errors, such as segments acting as satellites in multiple spans (2009*b*, p. 272) and “residual segments” (2009*b*, p. 273), that is, segments not participating in an RST span. Both of the errors identified by Thomas could also be found in the annotated corpus of this dissertation. Unfortunately, I will complement Thomas’ work with a description of two additional annotation errors. Fortunately, I also present a method for the effective verification of a GeM corpus.

The first case, which I term an ‘orphan span’, is created when the annotation fails to link one or more RST spans together (see Figure 5.3). The segments **s-3.01** and **s-3.02** are properly incorporated into the RST structure, unlike the segments **s-3.30**, **s-3.31** and **s-3.32** which fall outside of the entire RST structure. Orphan spans may occur anywhere in the RST annotation and they bear close resemblance to the “residual segments” described in Thomas (2009*b*, p. 273). In the DOT visualisation, the orphan spans are automatically placed on the right-hand side of the visualisation, where they can be easily located. The easiest way to locate and correct the error in the XML annotation is to search for the identifier preceding or following the orphan RST span’s identifier.

The second case — a ‘looping relation’ — occurs when the annotator inputs the wrong identifier, causing the RST span to refer to itself (see the SEQUENCE relation in Figure 5.4). Looping relations may be found in the visualisation by looking for relations that originate and terminate at the same span. Because the looping spans are properly incorporated into the annotation, instead of being placed on the right-hand side of the diagram, they are more difficult to locate than the orphan spans.

Finally, I describe how I implemented the overall verification procedure. I first used the RST visualisation script to create a large file in the DOT syntax, which included the visualisations for all RST structures in the annotated corpus (for more information on the visualisations, see Section 5.5.2). The file created in the process contained over 10000 lines of data, which also speaks for the need to

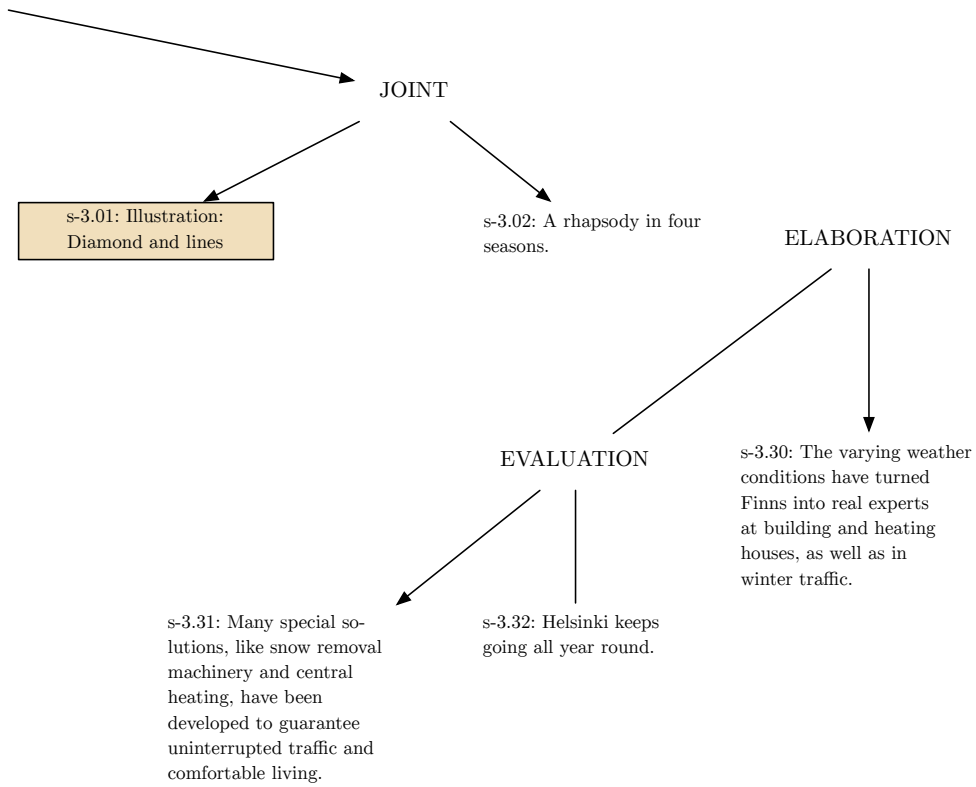


Figure 5.3: An orphan span in the annotation of *Welcome to Helsinki* (1998)

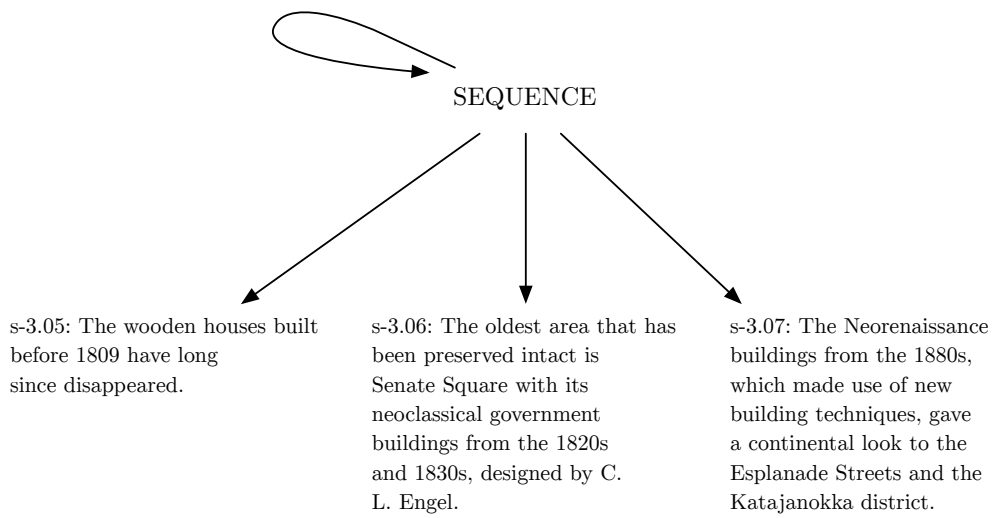


Figure 5.4: A looping relation in the annotation of *Welcome to Helsinki* (1998)



develop computer-assisted visualisations, as drawing these RST graphs by hand would be incredibly time-consuming.

I then used the *csplit* command in the OS X command line interface to split the file into separate files, using the curly brace } marking the end of each DOT graph to indicate the position where the file should be split. This resulted in individual files for each brochure in the annotated corpus. Finally, I used the command line interface to draw the DOT graphs for visual inspection.

I then proceeded to verify the annotation by visually identifying errors in the annotation. Upon the completion of the verification process, I performed a second pass on the data using the same procedure to check the implementation of the changes in the corpus. While I initially considered the corpus to be ‘frozen’ at this point, more errors were encountered during the development of queries, particularly in the form of missing values. It therefore appears that the creation of a multimodal corpus is an iterative process, and in this aspect similar to the creation of linguistic corpora (Biber 1993, p. 256).

## 5.5 Exploring the corpus

The successful application of the GeM model resulted in an XML-annotated corpus. As described above, the corpus was stored in multiple files using the XML markup language. In this section, I describe the technologies and techniques used for searching and exploring the XML-based corpus.

### 5.5.1 Query languages

Query languages are programming languages, which may be used to perform queries on databases. So far, several query languages have been proposed for interrogating GeM-based corpora. For instance, Bateman et al. (2002*b*, p. 7) pointed out that XSLT (Extensible Stylesheet Language Transformations) was cumbersome for complex queries, while the effectiveness of XQuery remained largely unexplored at the time. Thomas (2007, 2009*b*), in turn, combined XSLT and Perl to search and visualise GeM-annotated data. I decided, however, to work with query languages designed for XML, because there were potential benefits in using technologies belonging to the same family.

For this dissertation, the most important query languages were XPath (XML Path language) and XQuery, both of which have a World Wide Web Consortium (W3C) Recommendation.<sup>5</sup> A W3C Recommendation is a technical standard, which indicates that the query languages have been subjected to a review by the W3C organisation and the public. These query languages, their functions

---

<sup>5</sup><http://www.w3.org>

and capabilities have been described earlier in connection with the GeM model in Bateman et al. (2002a, pp. 28-29). I will now briefly present their application in this dissertation below.

Walmsley (2007, pp. 13-14) points out that XSLT and XQuery have many overlapping capabilities, while also identifying certain context-specific advantages of XQuery. XQuery is designed to select data from multiple files, whereas XSLT is generally used for transforming entire documents. This made XQuery a prime candidate for current work, because the GeM corpus stores its information into multiple files. XPath, in turn, is a query language for selecting and returning elements and attributes from XML documents, such as the GeM annotated corpus. To the advantage of applying XQuery to interrogate the corpus, XPath and XQuery overlap to a large degree and share the same data model and functions (Walmsley 2007, p. 13).

I will not introduce specific queries at this point, but simply wish to underline their importance in this dissertation. Because the queries selected and filtered information in the corpus, they needed to be carefully planned, in order to return the requested information. The queries, their structure and functions are described in connection with the analysis to bring out the interaction between the queries and the performed analyses.

Finally, particular attention in developing queries should be paid on the identifiers and their cross-referencing across the analytical layers. A query that retrieved the cross-referenced elements from each layer and organised them into table columns proved to be a useful tool for checking the annotation for missing identifiers. In this way, the missing values revealed if the query failed to account for certain types of annotation.

This concludes the description of verification measures. Next, I move to discuss the visualisation techniques used in this dissertation.

### **5.5.2 Visualisation techniques**

The final issue to be discussed is the visualisation of data. As a part of a growing trend, researchers working with multimodality (see e.g. Zappavigna 2010; O'Halloran et al. 2010, 2011; Smith et al. 2011), digital humanities and cultural studies (see e.g. Manovich 2012) have recently emphasised the importance of visualisation, particularly when dealing with complex and large data sets. O'Halloran (2009b, p. 6) points out that techniques of visualisation are a standard tool in natural sciences, because humans are better at interpreting complex data when it is presented visually:

The use of visualisation processes and computer graphics in mathematics and science for the interpretation of complex data sets relate to human capabilities of seeing visual patterns.

The benefits of visualisation also apply to the study of complex data in linguistic corpora. However, Siirtola et al. (2010) have noted that many of the current visualisation techniques have not been developed with linguists in mind, and consequently, many of them do not take the structure of language into consideration. Moreover, the requirements for visualising multimodal corpora may differ from the needs of monomodal corpora. Unlike linguistic corpora, multimodal corpora may combine very different types of information. In this sense, the GeM model is a prime example with its multiple layers for annotating the multimodal content, layout, visual features, navigation and rhetorical structure.

Previous work involving the GeM model have also included several types of visualisations. To begin with, Bateman et al. (2002a, p. 30) created GeM area models by transforming the XML annotation into HTML tables, using colour to indicate the distribution of verbal and visual content in the area model. Thomas (2009b) used two types of visualisations based on the Perl programming language: RST graphs to visualise the rhetorical structure of the artefact (2009b, p. 265) and “alignment graphs”, which show the segmentation of the analysed artefact (2009b, p. 271). The alignment graphs could be used to investigate how the base units are distributed and grouped as a part of other analytical layers.

The visualisations also play an important role in this dissertation. Like the previous work in Thomas (2009b) and Podlasov et al. (2012), I also used the DOT language as a part of the *Graphviz*<sup>6</sup> package for rendering the visualisations (Gansner and North 2000). In short, DOT is a language for drawing graphs: I used XQuery to transform the XML-based GeM annotation into DOT syntax. To clarify the issue, I will now briefly introduce the developed visualisations.

First, I rewrote the Perl-based RST graph script in Thomas (2009b, p. 265) in XQuery, while retaining its basic implementation due to its robustness, which I duly acknowledge here. The original script uses the RST spans and segments as DOT *nodes* and draws lines from the span’s nucleus and satellite segments towards the node as DOT *edges*. The content of the segments is drawn from the base layer and formatted with line breaks after 35 characters. In addition, the segments with visual content are marked using coloured boxes. Partial visualisations created using the rewritten script can be seen above in Figures 5.3 and 5.4.

Second, I modified the script to retrieve information from the layout layer to draw DOT *subgraphs*. The subgraphs are graphs within graphs, which can be used to represent different types of information stored in the layout layer. For instance, the entire RST graph could be organised into subgraphs according to the layout

---

<sup>6</sup><http://www.graphviz.org>

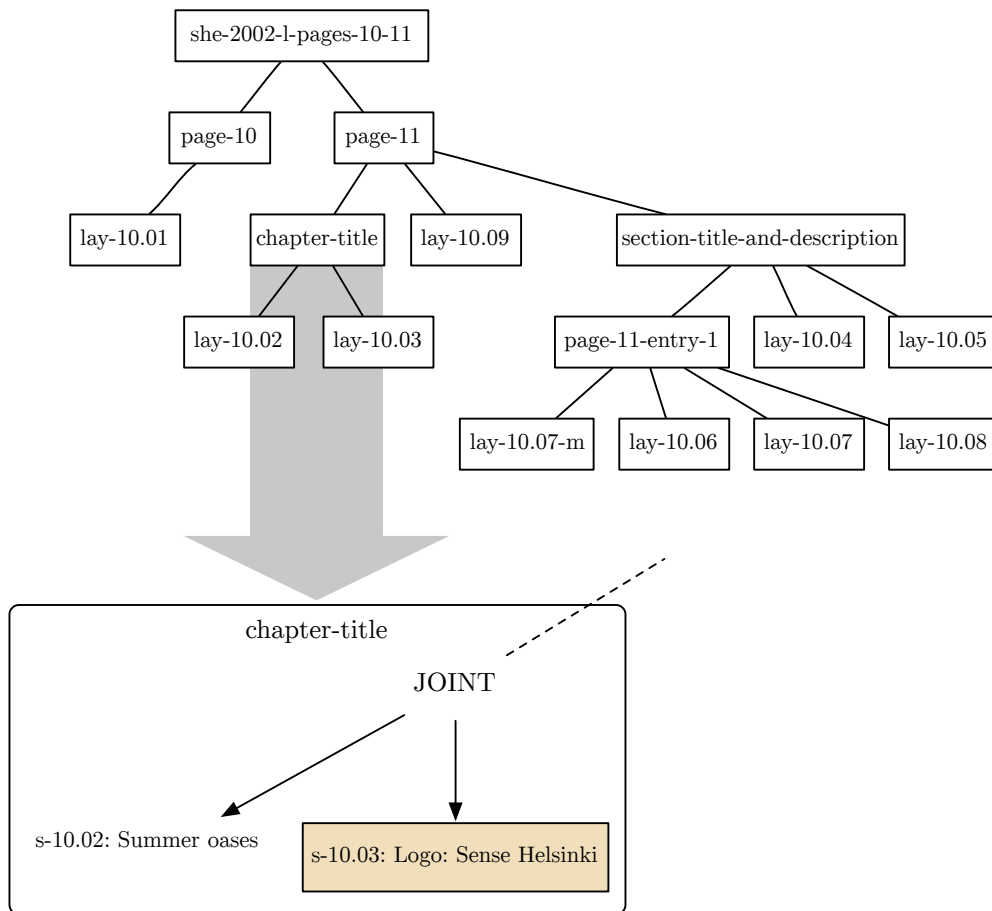


Figure 5.5: A key to the used visualisations

structure of the artefact, so that all RST segments located under the same parent node in the layout structure are drawn inside the same subgraph (see Figure 5.5). Alternatively, the RST segments could be reorganised into subgraphs according to their placement in the area model. In this way, the modified script considerably enhanced the dissertation’s capability to observe the interaction between the rhetorical and the layout layer (Hiippala 2013).

I will now explain how the visualisations should be interpreted using the example shown in Figure 5.5. The top half of the diagram is occupied by a familiar representation of the layout structure: a tree diagram, which represents the hierarchical organisation of the content. The bottom half of the diagram shows a partial RST diagram with two segments: **s-10.02** and **s-10.03**. At this point, attention should be paid to the bounding box around the two segments, which is entitled *chapter-title*. It is also the name of the layout chunk that can be found in the tree diagram above. This shows the RST script in action: bounding boxes

are drawn around the RST segments that belong to the same layout chunk in the layout structure.

At this point, it is useful also to recall the schema used in RST diagrams. In the representations of RST structure, the lines indicate the role of the segments in the RST structure. A line with an arrow points towards the nucleus. For example, in the case of Figure 5.5, there are two nuclei (**s-10.02** and **s-10.03**), because `JOINT` is a multinuclear rhetorical relation. In Figures 5.3 and 5.4, the lines without an arrow indicate the segments acting as a satellites. Dashed lines, such as those shown in Figure 5.5, indicate `TITLE` segments. For a complete description of nuclei and satellites in RST, see Section 3.2.3.

Before the conclusion, a final word needs to be said about developing visualisations: an XML-to-DOT transformation has to accommodate all possible annotations allowed by the DTDs (see Section 5.3.2). This requires a careful, iterative development process with sufficient testing, especially in handling the identifiers and their cross-referencing across the analytical layers. To conclude, combined with carefully planned and written queries, the XML-based technologies provided new means for storing, searching and visualising the data. As Bateman et al. (2002a, p. 29) have pointed out, the interests that drive the development of XML-based technologies are primarily economic. And because the GeM model uses XML, it also benefits from state-of-the-art technologies developed for handling XML data.

This concludes the description of the data and the application of the GeM model. Although this description may have appeared as highly technical, the real potential of the GeM model becomes evident in the following chapters, which model the structure of a multimodal artefact on the basis of the GeM-annotated corpus.

# Chapter 6

## The medium and its characteristics

In the following chapters, I will show how the theoretical framework developed in Chapters 2, 3 and 4 may be used to model the structure of a multimodal artefact. In this task, the two data sets described in Section 5.1 will play a central role. Their role may be highlighted using a quote from Halliday (1992, p. 356):

By bringing [a text] under grammatical attention we display its properties as an instance; this shows how it means, since every feature it displays is located in the context of its alternatives. These agnate forms, representing what might have been meant but was not, constitute an infinity of shadow texts against which the one in focus achieves its reality.

The annotated corpus provides a context of alternatives, which reflects the semiotic choices made in the tourist brochures. Using the annotated corpus, we can map the choices that are preferred over alternative choices, and these preferred choices may reveal the characteristics of the tourist brochure as a multimodal artefact. Although the annotated corpus equals only a drop in the ocean, particularly if contrasted with the size of the currently emerging data sets (cf. e.g. Michel et al. 2011; Manovich 2012), it is necessary to tread this uncharted territory carefully, as our experience with multimodal corpora is still limited.

Quantifying and analysing the raw GeM data alone cannot pinpoint which decisive semiotic choices enable us to recognise the analysed artefact as a tourist brochure (see e.g. the results in Hiippala 2012*b*). Consequently, the only way to move forward is to bring the theoretical framework — and particularly the notion of semiotic modes — to bear on the data. Equally important is to open new analytical perspectives by combining information from the multiple analytical layers of the annotated corpus (cf. Hiippala 2013). Furthermore, the analysis needs

to proceed in a logical order: from the medium to the semiotic modes and to their interpretation, so that the contribution of each factor to the structure of a multimodal artefact may be identified.

What all of the above means in practice will be laid out in the successive chapters, following the model of an artefact's multimodal structure described in Section 4.5. This chapter begins with a discussion of a more abstract concept, in this case, the notion of the brochure as a *medium*. I shall then elaborate on various aspects of this medium and its characteristics. To do so, Section 6.1 outlines some general features of the medium, continued by a perspective on advertising in Section 6.2. Finally, Section 6.3 addresses the medium's visual qualities and certain aspects of its production.

## 6.1 Why the notion of a medium matters

Defining a medium (or its plural form *media*) is challenging due to the widespread use of the concept in both academia and everyday life. To exemplify, print media, digital media and social media are some of the frequently used concepts. For the sake of lending more analytical value to the concept, I shall now consider why the medium matters when describing the structure of a multimodal artefact.

To begin with, Bateman considers medium “a historically stabilised site for the deployment of some selection of semiotic modes for the achievement of varied communicative purposes” (forthcoming, p. 12). In this view, examples of media that fall under the umbrella of print media could include — for instance — newspapers, books, magazines, brochures and leaflets. Each of the aforementioned media may prefer a selection of certain semiotic modes. Depending on the kind of genre *realised in the medium*, specific choices within the semiotic modes may be favoured. As a part of a rhetorical strategy, these choices are considered effective for the genre and its communicative goals (cf. Hiippala in press).

At this point, the *genre* of a tourist brochure, which was described in Section 4.3.1, needs to be clearly contrasted with the *medium* of a brochure. The main question is: does the medium provide something besides the material substrate to realise the genre? Bateman (forthcoming, p. 11) argues that features such as page numbers, text spacing and paragraphing, the conscious use of empty space for margins, etc. do *not* contribute to the multimodal genre structure, neither does their presence, absence, form or placement.

These highly conventional features are independent of any genre, because they arise as a result of the process of pagination and layouting — the placement and organisation of content on pages — and can appear in any genre realised using the medium. Consider, for instance, this very page as a part of the genre of a doctoral dissertation (see e.g. Paltridge 2002; Bunton 2005), which is realised in

the medium of a book. The genre of a dissertation, in turn, is realised using the semiotic mode of paragraphed text-flow, surrounded by margins and accompanied by a page number. Yet the paragraphs, margins and page numbers are not here to advance the argument of this dissertation: they increase legibility and help to navigate the manuscript.

As a part of the multimodal artefact, these features perform an equally important task as the content, whose organisation they support. Collapsing their contribution into a single, unified structure would do justice to neither the genre nor the medium. The distinct contributions of the medium, the semiotic modes and the genre come together in a multimodal artefact — the target of the current investigation — and for this reason, their contribution needs to be taken apart clearly in the analysis (Bateman forthcoming, p. 13).

Now, in contrast to the common features of the medium of a book, what can be said about the brochure as a medium? In the following section, I shall attempt to describe the medium of a brochure and to establish its properties. This also sets the stage for the discussion of the semiotic modes and their relation to the medium of a brochure.

### 6.1.1 The brochure as a medium

In previous research, the tourist brochures have been described as “polysemic” constructs and texts (Edelheim 2007; Valdeón 2009), as “communicative acts” (Yui Ling Ip 2008), as texts belonging to the “institutional tourist genre” (Francesconi 2011), and as a genre of print media (Hiippala 2007, 2012*b*). Naturally, these definitions reflect the analytical interests of the particular studies. However, none of the previous research has considered the brochure as a medium. Therefore, I shall now explore the brochure as a medium, which carries the semiotic modes that are used to realise the *genre* of a tourist brochure.

Initially, the broad criteria for describing the medium of a brochure need to be set out. As Bateman points out, “no meaning can be realised in a medium”, because “meaning can only be *realised* in a semiotic mode *participating* in a medium” (forthcoming, p. 12). Therefore, the medium of a brochure acts as a carrier of the semiotic modes, which realise the genre of a tourist brochure. Consequently, the umbrella term of print media may be loosely described according to the semiotic modes they can deploy. Due to the same material substrate, print media can also deploy the same semiotic modes.

However, the actual difference between the media is likely to emerge in the *configuration* of the semiotic modes and their use. Consider, for instance, the remarkable differences in the use and configuration of text-flow in the medium of a newspaper and the medium of a book. Capturing the structure and the function of the semiotic modes is precisely what the GeM model is used for in



this dissertation and for this reason, the results may also inform us about the properties of the used medium to a very high degree of detail. To achieve such a detailed view, it is necessary to begin by considering the basic properties of the medium, which leads us to the issue of materiality.

As said, the medium of brochure is defined by its material substrate (see Section 3.3.1). At the core of the medium is the material substrate of a printed page, which provides a range of semiotic modes in their *static* form while simultaneously constraining the deployment of their *dynamic* forms. One simply cannot use the brochure to realise a film, because the printed page cannot carry the semiotic mode of dynamic image-flow. Furthermore, similar to other forms of print media, such as newspapers and books, the physical space available for the semiotic modes in the medium of a brochure may differ. In the annotated corpus, the size of the brochures ranged between 2128 cm<sup>2</sup> for the largest brochure (*HDB* 1972) and 295 cm<sup>2</sup> for the smallest (*SHE* 2002).<sup>1</sup>

To explore the issue further, I shall now consider two relevant properties of the medium: the method of binding and the fold geometry. The method of binding is concerned with how the pages are joined together. The fold geometry, in turn, describes how the pages may be folded to manipulate the size of the artefact.

## 6.1.2 Brochures and leaflets

A quick look at the entire data set revealed a difference that could be used to divide the data into two different categories. This difference was the method of binding, that is, whether staples were used to join the pages or not. The method of binding has consequences for artefacts that rely on a “page metaphor” to organise the content (Bateman 2008, p. 9): whereas a staple-bound brochure can add up to four content pages by adding a single sheet of paper, a leaflet without staples may only expand the available content space by adding a folding point. A folding point, which I will define shortly, provides more content space without increasing the space needed to store the brochure. The small size is important for the consumption of the tourist brochures, as they are meant to be carried around easily (Hiippala 2007, p. 11).

To investigate this issue, each instance in the entire data set was coded according to its method of binding: the artefacts were either bound using staples (n = 44) or not (n = 45). This resulted in two categories: *brochures* for staple-bound artefacts and *leaflets* for artefacts without staples. The categories of brochures and leaflets were then compared using two continuous variables: (1) the number of content pages and (2) the number of folds per content page.

---

<sup>1</sup>The identifiers given here and afterwards refer to the identifiers in Table 5.2 in Section 5.3.

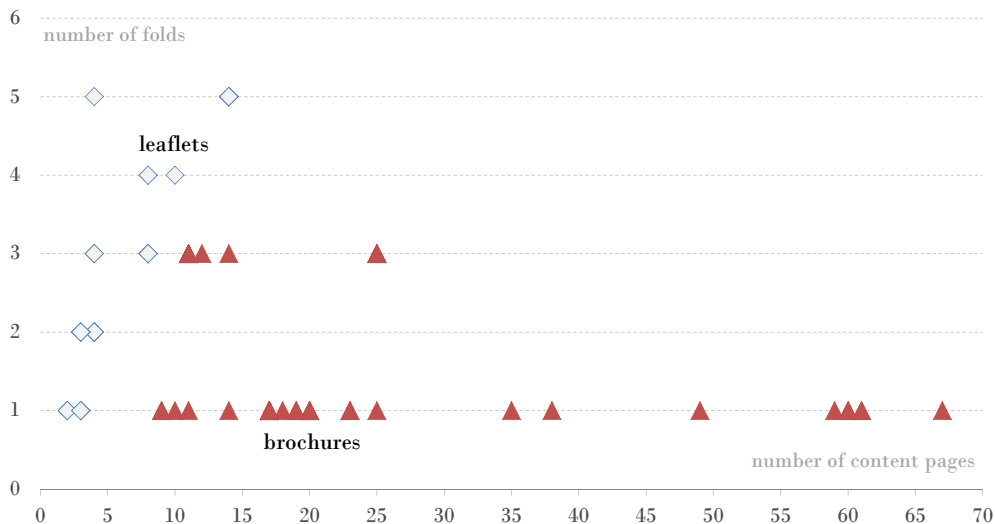


Figure 6.1: The number of folds and content pages in brochures and leaflets

At this point, it is necessary to provide the definitions related to variables (1) and (2). First of all, a content page is delimited by a fold or the edge of a page in any direction. In this case, a fold is considered a point where the physical page is folded to reduce the artefact’s overall size. Moreover, a content page has to realise some stage of the genre, which were defined in Section 4.3.1. Pages that contained only navigational elements, such as indices, and the pages arising from the medium itself, such as covers and leftover pages, were excluded from the count. The resulting distribution of the brochures and leaflets ( $n = 89$ ) is shown in the scatter plot in Figure 6.1.

As Figure 6.1 shows, the leaflets grow their content space by increasing the number of folding points. The brochures, in contrast, increase the available space by adding more pages. Interestingly, in the region between 10 and 30 pages, a number of the brochures ( $n = 13$ ) also use the fold geometry to increase their content space. In some cases, an additional folding point is used to reduce the overall size of the artefact.<sup>2</sup> In other cases (*SHF* 1987), a folding point is added to enable the inclusion of maps. This is done by embedding a smaller, bound brochure within larger cover pages, which contain a fold-out map. It should be noted, however, that exploiting the fold geometry is not a new phenomenon: the expansion of content space by adding folding points can also be found in other print media, such as magazines (see e.g. O’Halloran and Lim 2009).

In more general terms, Figure 6.1 shows that staple-bound brochures include more content pages. The mean values and standard deviation for the number of

<sup>2</sup>*HFT* 1976, 1984, 1988; *TGH* 1980.

Table 6.1: The number of content pages in brochures and leaflets

	Brochures	Leaflets
Mean	24.6	6.8
Standard deviation	16.7	4.5
P-value	<0.01	

content pages in both brochures and leaflets are given in Table 6.1. In both cases, the standard deviation is relatively high. This means that the instances of data, which reflect the number of content pages, are spread out over a range of values. In terms of the mean value, the number of content pages in the brochures is three times as high as the number of content pages in the leaflets. This suggests that a staple-bound binding makes it easier to increase the space available for content, because a single sheet of paper can add up to four content pages to a brochure.

Concerning the validity of the categories in Table 6.1, a two-tailed t-test showed a statistically significant difference between the number of pages in the brochures and leaflets ( $P < 0.01$ ). Following the convention in humanities and social sciences, a P-value of less than 0.05 is considered a cut-off point (Levon 2010, p. 71). It is therefore reasonable to believe that the two categories differ in terms of the number of content pages.

What does this observation imply? From the perspective of the GeM model, a production and consumption constraint may explain the difference between brochures and leaflets (Bateman 2008, p. 16). In simple terms, the planned number of content pages seems to determine whether the artefact should be either staple-bound or folded. Including 67 content pages — the largest number of content pages in a brochure — into a folded leaflet would present the artefact’s user with serious challenges: accommodating this number of content pages would require nine horizontal and nine vertical folding points. Anyone familiar with large, folded maps will also know the difficulty of folding them back together.

In contrast, a staple-bound brochure can easily include the same number of pages (HVG 2008). However, Valdeón has observed in a study of Spanish tourism texts that “the distribution of information in the brochures and leaflets differs considerably” (2009, p. 26). Due to the linguistic focus of Valdeón’s study, he did not pursue this issue further.<sup>3</sup> For the multimodal analyst, the follow-up question is naturally whether the choice of the material substrate and the method of binding also affects the configuration of the semiotic modes. In short, the entire data set

<sup>3</sup>In personal communication on 25 January 2013, Valdeón stated that the main difference between the brochures and the leaflets was their layout. Whereas the leaflets were of the “folding and unfolding type” and had to be turned around to locate different information, the booklets were “less beautiful to look at” but easier to use.

revealed two possible variants of the medium: the brochure and the leaflet. The annotated corpus, in turn, can be used to study the multimodal structure of the brochures and leaflets, which I hereby term *medial variants*, adopting the term proposed by Stöckl (2004), albeit in a different context (see Figure 2.3). I will return to the potential differences of the medial variants in Section 8.5.

So far, I have covered certain basic properties of the brochure as a medium. These properties consisted of standard medium-related features, such as paragraphs, margins and page numbering, which result from the process of pagination. I also considered the method of binding as a key feature that distinguishes the staple-bound brochures from the folded leaflets. The method of binding also determines the amount of content that can be included in the artefact realised using the medium. In the following section, I take a step towards the content by considering the role of advertisements, which may co-exist in the medium alongside the genre of a tourist brochure.

## 6.2 The relationship between a medium and advertising

In this section, I outline an approach to advertising and media within the developed framework. I argue that in the model of multimodal structure, the genre and the advertisements need to be separated due to their structural and functional differences. They both contribute to the artefact in terms of content and structure, but their contribution should not be collapsed into a single entity. To explain my view, I begin with a generic description of advertising from a multimodal perspective.

### 6.2.1 A multimodal perspective to advertising

In the contemporary consumer society, advertising may be broadly defined as an activity that encourages the consumption of goods, services and information. At the same time, advertising is a multifaceted phenomenon that may be approached from purely economic standpoints, or alternatively, it may be described as a series of complex semiotic and psychological operations to persuade both individuals and groups. By using the word ‘complex’, particularly in connection with the term ‘semiotic’, I wish to underline that advertising is not a simple, straightforward process of communication between the advertiser and the consumer. Instead, the persuasive goals are often pursued through indirect discourse (cf. Crook 2004; Catenaccio 2008), across a variety of media (cf. Lemke 2009; White 2010, 2012), and most importantly, by using multiple semiotic modes (see e.g. Cheong 2004;

Hopearuoho and Ventola 2009; Tan 2009; Feng 2011; Martínez Lirola and Chovanec 2012).

Advertising has been also studied from a semiotic perspective outside the field of mainstream multimodal research (see e.g. Forceville 1996; Beasley and Danesi 2002; Van Mulken et al. 2010; Martínez-Camino and Pérez-Saiz 2012; Lagerwerf et al. 2012). The contribution of multimodal research, in turn, may be said to be an increased understanding of the semiotics of advertising on different levels of abstraction. For example, Royce (1998), Cheong (2004) and Feng (2011) have provided detailed descriptions of meaning-making in printed advertisements. On a more abstract level, O'Halloran and Lim (2009) and Berazhny (2012) have contributed broad descriptions of the advertisements in the medium of a magazine. Again, the notion of a medium brings us to the issue at hand: our knowledge of advertising within the medium of a brochure is limited, which provides a challenge for modelling the structure of a multimodal *artefact*.

In order to meet this challenge, I propose that within a model of artefact structure, advertising should be considered to be a medium-based phenomenon. This distinguishes the contribution of advertising from the structure of a multimodal artefact and separates the advertisements from the genres that are carried in the deployed medium. However, I want to emphasise that this does not rule out studying specific kinds of advertisements as a genre in other contexts (see e.g. car advertisements in Hopearuoho and Ventola 2009; Feng 2011). For this dissertation, however, separating the genre and the advertisements is necessary, because advertising is driven by decisions related to marketing strategies that attempt to reach the potential consumers who also have an interest in the deployed genre. The semiotic choices made in the advertisements are of secondary concern and largely differ from the choices made in the genre: I will now elaborate my view below.

The advertisements have the potential to be realised in a medium, if the combination of the medium and the genres affords advertising. This constraint needs to be set out clearly, because certain combinations do not afford advertising. Consider, for instance, factual and fictive genres realised in the medium of a book. For example, this dissertation does not have classified advertisements, except for scholarly purposes (cf. Figure 6.3). This does not, however, prevent all forms of advertising: it may also be implicit and 'embedded', such as in the case of product placement (Berazhny 2008). Alternatively, advertising may be explicit and take up parts of the layout space (cf. O'Halloran and Lim 2009; Berazhny 2012). In the following discussion, I will focus on these explicit forms of advertising.

To exemplify the phenomenon, let us first consider advertising in the *medium* of a printed newspaper. For instance, Lemke (2005, p. 46) suggests that advertisements constitute a multimodal genre in the newspapers. He writes:

The printed advertisement is a genre that we might argue obligatorily includes an image as well as a text that has a full organizational structure of its own (as opposed to being merely a labelling or gloss on an image).

We already possess a considerable amount of knowledge about the multimodal characteristics of both printed and digital newspapers (see e.g. Ventola 2005; Knox 2007; Bateman et al. 2007; Caple 2009a; Caple and Bednarek 2010) and their perception (Holsanova and Holmqvist 2006). If our knowledge of newspapers is contrasted with our knowledge of multimodality in advertising, it should become clear — as Lemke argued above — that we are dealing with two different types of semiotic phenomena: the genre of a newspaper and the genre of newspaper advertisements, both of which have different communicative goals. Most importantly, both have their own multimodal organisation in terms of structure. This is precisely why their contribution needs to be separated when modelling the structure of a multimodal artefact, because collapsing these contributions will result in the loss of analytical focus, if two distinct forms of semiosis are treated as one.

A further example may be drawn from a study of in-flight magazines as a mixed genre, which combines both editorial and advertising content as a “blend of travel brochure, lifestyle magazine, corporate catalogue and information leaflet” (Thurlow and Jaworski 2003, p. 584). In relation to multimodality, Thurlow and Jaworski suggest that the magazines are “spatially integrated texts” that “appear to be heterogeneous in drawing from different publishing genres, while at the same time being extremely consistent in how they do that” (2003, p. 585). Curiously, Thurlow and Jaworski (2003, p. 585) consider the mix of genres and the consistent multimodal organisation as *paradoxal*: let us now evaluate this statement in more detail.

It may be argued that the above statement by Thurlow and Jaworski results from a failure to account for the complexity of the in-flight magazine as a multimodal artefact. This, in turn, is a direct consequence of collapsing the different contributing factors into a single notion of genre. In this case, a clearly formulated notion of the magazine as a medium would have provided a more informed perspective into what exactly comes together in an in-flight magazine in terms of the genres and the semiotic modes. As Bateman (forthcoming, p. 12) points out, a *medium* such as a magazine may carry an unlimited number of genres, but the medium is simultaneously constrained by production, consumption and genre constraints, which also result in the kind of “consistent” structuring of meaning, as described above by Thurlow and Jaworski (2003, p. 585).

Finally, I will draw on yet another example from outside of the current field of investigation. The co-existence of advertisements is also a common feature in connection with dynamic, audiovisual discourse. In connection with filmic (Bateman

and Schmidt 2012) and telecinematic (Piazza et al. 2011) discourse, the advertisements which precede, interrupt or follow the unfolding discourse are not considered as a part of the analysed filmic or telecinematic artefact.<sup>4</sup> This view is precisely the same I advocate for studying the structure of printed multimodal artefacts, that is, the clear separation of the genres and the advertisements. With the description of advertising now complete, I shall proceed to investigate the occurrence of advertisements in the entire data set.

## 6.2.2 Advertising in the tourist brochures

In order to establish a perspective into advertising in the tourist brochures, I analysed the entire data set to determine whether the instances of data included layout space for advertisements or not. For print media, there are two established categories for describing advertisements. The first category is that of *display* advertisements, which occupy an entire page. Display advertisements have been frequently analysed in previous multimodal research (Royce 1998; Cheong 2004; O'Halloran and Lim 2009; Feng 2011). The second category consists of *classified* advertisements, which are smaller in size and organised according to the products and services they offer. Figure 6.3 shows classified advertisements in the *Weekend Delight in Helsinki* brochure, published in 2006, which advertise a range of different products, using an equally wide range of graphic, typographic and layout choices (cf. Lemke 2005, p. 46).

Because the tourist brochures themselves constitute a form of marketing and advertising (Molina and Esteban 2006), the following criteria were used to distinguish between display and classified advertisements in the data:

1. The advertisements occupy their own layout area and can be clearly distinguished from the main content of the tourist brochures.<sup>5</sup>
2. The advertised service or product is not directly affiliated with the publisher of the tourist brochure.

As Figure 6.2 shows, the classified advertisements appear in the data only after the year 2000 ( $n = 89$ ). Moreover, all of the brochures with classified advertisements are staple-bound and most have a high number of content pages.<sup>6</sup> This should not come as a surprise, as these brochures have more content space available: the notion of “selling media” — that is, advertising space — is nowadays an established field of business (see e.g. Warner 2009). The benefits are obvious,

---

<sup>4</sup>In this connection, it is also important to note the role of product placement in film as a another means of ‘colonising’ a medium (see e.g. Scolari 2009).

<sup>5</sup>In addition, advertisements typically need to be recognisable and adhere to legal requirements. For a case study of this issue in Russia, see Berazhny (2012).

<sup>6</sup>*SHE* 2002; *HYW* 2003, 2005, 2006; *WDH* 2006; *HVG* 2008.

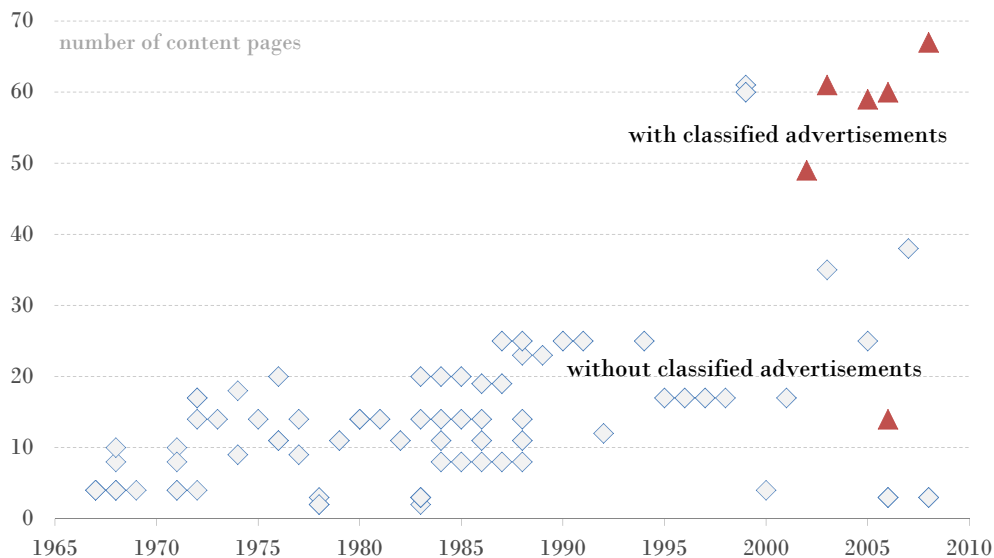


Figure 6.2: The distribution of advertisements in the tourist brochures

because selling advertising space may compensate publishing costs and enables the advertisers to tap into the lucrative tourist market, valued worldwide at 740 billion euro in 2011 (World Tourism Organization 2012).

In addition, the inclusion of classified advertisements may broaden the overall function of the tourist brochures. Several advertisements in Figure 6.3 offer a discount to the holder of the brochure, thereby expanding the brochure's functions. The brochure does not only inform and advertise, but also directly encourages the consumption of goods and services by simultaneously functioning as a discount coupon. The brochure acknowledges this explicitly:

This brochure has been designed to help you enjoy Helsinki without having to spend unnecessary amount of time planning your next move. Besides useful information about places to go and things to do, you will find plenty of valuable offers from local shops, restaurants and tour operators. Helsinki Expert coupons should be presented for discounts when purchasing. Other discounts you can claim by simply presenting this brochure at the point of purchase.<sup>7</sup>

This observation may be of interest for marketing researchers, but the additional function of the brochures does not make a difference in the structure of the artefact, because this function arises from the classified advertisements, whose structure is not described within this dissertation.

<sup>7</sup> WDH 2006: u-1.12 to u-1.15.





Figure 6.3: Classified advertisements in *Weekend Delight in Helsinki* (2006)

To conclude, given the relatively low number of brochures with advertisements ( $n = 6$ ), the main concern of this dissertation should continue to be the multimodal structure of the tourist brochure as an artefact. With the basic characteristics of the medium thus established, I now turn to describe the semiotic modes available in the medium. Obviously, these semiotic modes comprise the written language and various types of images. In particular, previous research has emphasised the role of the visual in the tourist brochures (see e.g. Scarles 2004; Garrod 2009; Jokela 2011). For instance, Scarles (2004, p. 45) observes that:

Brochures are highly visual in nature and rely upon images and photographs to sell destinations as tourist places become immersed in the language of visuals.

In the following section, I investigate whether the annotated corpus agrees with this observation. The aim is to uncover the distribution of verbal and visual content in the data, which has been assumed to have favoured the visual in the recent years.

### 6.3 Aspects of production

This section considers the tourist brochures from two perspectives. First, Section 6.3.1 explores the so-called ‘visual turn’ on the basis of the annotated corpus. The following Section 6.3.2, in turn, discusses the impact of technological development on the production of the tourist brochures.

### 6.3.1 The ‘visual turn’ in the tourist brochures

The implications of the so-called ‘visual turn’ in communication have been discussed in various fields of study (see e.g. Kress 1998; Livingstone 2004; Duncum 2004; Mitchell 2005). In multimodal research, both theoretical (Iedema 2003) and methodological (Bateman et al. 2004; Parodi 2010) concerns have been raised in connection with the visual turn. Moreover, specific observations have been made about artefacts that are functionally close to the tourist brochures. As Kong (2006, p. 222) writes:

Travel guides exhibit a range of styles, from content that is extremely packed with words without any pictures to content that is fully illustrated with images and pictures. This reflects the history of the genre — texts produced more than 10 years ago tend to be more word-based, whereas those produced in the last 5 years are usually illustrated with more pictures and images, although the degree of integration between words and images can vary considerably.

Kong’s observation raises a question relevant for this dissertation: has there been a similar development in the tourist brochures?

To investigate this issue in the annotated corpus, I performed two queries on the occurrence of graphic elements. The first data set returned by the query, represented by the line in Figure 6.4, consisted of the base units with *alt* attributes, which indicate visual content (see Section 5.2.1). Drawing on the layout layer, the second set of data consisted of the total size of the *graphics* elements in the realisation information, which had any of the following *type* attributes: photo, illustration, and map. The second data set is represented by the bars in Figure 6.4. I shall now discuss both queries in greater detail.

To begin with, it should be noted that the average percentage of visual base units in the data is relatively low (17.4%). The low average can be explained by the GeM annotation schema. A photograph constitutes a single base unit in the base layer, whereas the minimal unit of analysis for language is limited to an orthographic sentence or a sentence fragment (see Section 3.2.1). If the number of visual layout units were compared to that of verbal layout units — that is, entire paragraphs — the mean value for the percentage of graphic elements would have been higher. Nevertheless, a broad survey of the distribution of visual and verbal content can already tell us many useful things about the multimodal structure of the tourist brochures, as I will show below.

Note the range of variation in the number of visual base units in Figure 6.4 and that in several instances, the brochures contain only linguistic base units.<sup>8</sup>

---

<sup>8</sup>*MEH* 1967; *CEH* 1967; *WTZ* 1978; *WTH* 1983; *SMH* 1982; *HWY* 1999; *BNO* 2006.

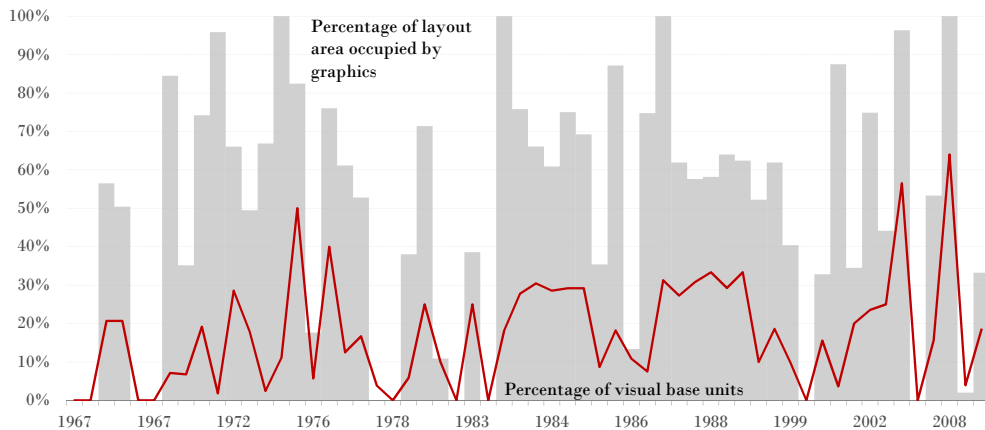


Figure 6.4: Graphic elements in the annotated corpus

Furthermore, in the light of the proposed visual turn, these non-visual instances appear throughout the years represented in the data (1967-2008), shown on the horizontal axis in Figure 6.4. On the one hand, among the “range of styles” (Kong 2006, p. 222) for the tourist brochures, there also exists a configuration that prefers purely linguistic communication. On the other hand, Figure 6.4 also shows spikes where the percentage of visual base units exceeds 50% (1976, 2006, 2008). This appears to represent an alternative configuration, which relies more on visual communication, as opposed to its verbal counterpart. As I will show, the spikes and dips in Figure 6.4 correspond to specific configurations of the semiotic modes in the tourist brochures, which are described later in Chapter 8.

To conclude, the base layer can only quantify some very basic properties of the tourist brochures. Among the two opposites, the spikes and the dips in Figure 6.4, there are configurations that seem to divide their communicative tasks between the verbal and visual modes without preferring neither. It is precisely here where the attention needs to be focused, if we aim to study what Kong (2006, p. 222) calls the “degree of integration” between words and images. I have previously argued that to describe the integration of the semiotic modes, it is necessary to combine information from multiple sources: this also applies to the investigation of the visual turn in print media (cf. Mitchell 2005).

Consequently, the bars in Figure 6.4 combine information from two different components of the layout layer, in order to evaluate the effect of the postulated visual turn on the layout space and its use. The information was retrieved from the area model and the realisation information components of the layout layer. This was done using a query, which first calculated the total amount of available layout area and then proceeded to calculate the size of each graphic element present in

the layout. The result shows that the graphic elements take up a significant part of the available layout space in the annotated corpus: the average amount of space taken up by graphic elements was 50.5%.<sup>9</sup>

As I noted above, the calculated size for the graphic elements included only the following element types: photo, illustration, and map. This means that certain two-dimensional elements, such as demarcating lines and in-line icons were excluded from the count, thus focusing on the aforementioned graphic elements that are known to do communicative work (see e.g. MacEachren 1995; Machin 2004; Boeriis and Holsanova 2012). As Figure 6.4 shows, images have been present in the tourist brochures throughout the period of time represented by the data. Therefore, it is not surprising why it has been suggested that the graphic elements play an important role in the tourist brochures, particularly in the formation of an image of the destination (cf. Molina and Esteban 2006).

What comes to the extensively discussed visual turn, it appears that the tourist brochures have been fairly visual all along, at least according to the annotated corpus. It should be kept in mind, however, that the visual turn does not necessarily manifest itself as a growing proportion of graphic elements or the space they use (Hiippala 2013, p. 466). Instead, if a visual turn has taken place, it may also manifest itself as novel configurations of the semiotic modes: I will investigate this possibility in Chapter 8.

In conclusion, I want to refer back to the hypothesised model of change for multimodal artefacts and genre, shown in Figure 3.1. On the basis of the model, I propose that the postulated visual turn may be largely based on everyday observations of *semogenic processes*, particularly in terms of phylogenesis — the long-term expansion and change that affects the semiotic modes (see Section 4.3.2.1). Humans appear to be sensitive to phylogenesis, which enables us to estimate when a multimodal artefact was produced. This kind of phylogenetic sensitivity may be a part what has been described as “document” or “functional” literacy (see Cohen and Snowden 2008; Waller 2012), and casual statements about certain multimodal artefacts being ‘retro’, ‘vintage’ or ‘old-fashioned’ may reflect this sensitivity (cf. Kress and van Leeuwen 2001, pp. 72-73).

Whether the phylogenetic sensitivity also enables us to accurately describe the specific characteristics — such as the degree of visuality in multimodal artefacts — without analytical training is a completely different issue. I thus suggest that a more neutral perspective to the phenomenon can be achieved through multimodal corpora. As the annotated corpus shows, the tourist brochures have been relatively visual all along: this calls for a systematic investigation of the so-called visual turn,

---

<sup>9</sup>In some cases, the value exceeded 100%: this results from a limitation of the GeM model in measuring and annotating non-geometric shapes (see Section 5.2.2.1). These values have been cut off at 100%.

in order to avoid working with mere assumptions. It is important to raise this point, because the possibly unfounded notions of a visual turn may also influence the society at large through educational policy: I shall explicate my argument below.

A proper investigation of the visual turn should thus involve computer-based approaches and larger data sets. Again, the work of Michel et al. (2011) is a promising lead: the data includes a wealth of media that use both language and image, for instance, various types of printed magazines. After circumscribing a corpus consisting of a specific medium and retrieving the data, it would be possible to calculate the distribution and proportions of verbal and visual content, as I have done in Figure 6.4 above. Turning the manual process into an automatic one, however, will require the input of specialists in computer vision (see e.g. Okun et al. 1999). Combined with appropriate methods of multimodal analysis — such as those developed in this dissertation — a model of the multimodal artefact could be used as the basis for identifying the preferred multimodal structures.

Consequently, the research described above could inform us whether (1) the communication has indeed shifted towards the visual in the studied media, and (2) if this shift has also affected the multimodal structure of the artefacts realised in the medium, thus answering the call for writing the history of “visual language” (Kostelnick and Hassett 2003, p. 231). This kind of knowledge could prove valuable for informed policy-making in the various fields of education (see e.g. Hobbs 1998; Freedman and Stuhr 2004) and considerably sharpen our understanding of what goes on in a multimodal artefact. This knowledge could support a consideration of what needs to be taught and how to reach a level of functional literacy required in the world today, encouraging questions such as:

- Do we need to teach the critical analysis of the content or should we focus on teaching how communication is structured multimodally?
- Should we emphasise the interpretation of visual content or the combinations of both verbal and visual content?
- If all aspects presented in the questions above are relevant, what is the most efficient way of teaching them simultaneously?

Providing the answers to these questions will most likely require input from multiple disciplines, but most importantly, all of the answers will need to be informed by empirical research. Moreover, if we are indeed advancing into an age defined by multimodal communication, the effects of technology need to be considered as well. For this reason, in the following section I discuss the impact of technological development on the structure of a multimodal artefact. To keep the discussion within the scope of this dissertation, I shall focus exclusively on the aspects of *production*, as opposed to the technologically-mediated consumption of multimodal

artefacts (for research in this area, see e.g. Leander and Vasudevan 2009; White 2010, 2012; O’Halloran and Smith 2012).

### 6.3.2 The impact of technological development

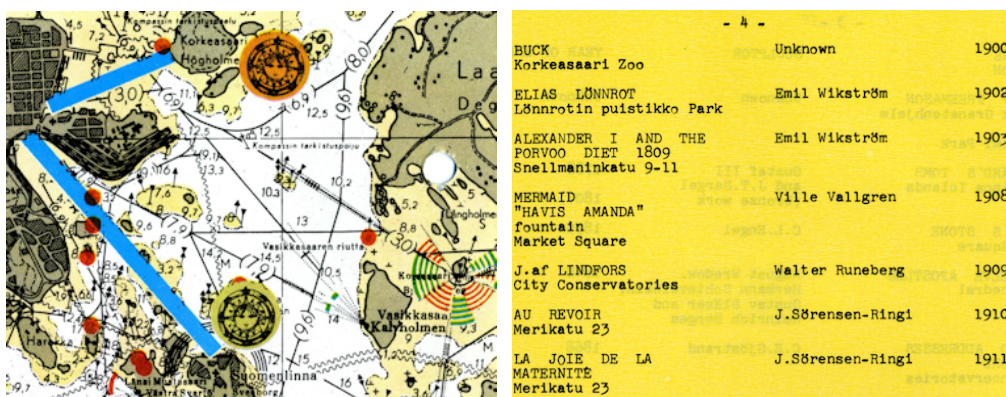
Technological development has transformed the design and creation of printed artefacts. This is not an understatement, considering the changes that have taken place in the production of the tourist brochures between 1967 and 2008. In this section, I highlight some aspects of producing print media artefacts that have been somewhat neglected up to this point. For example, Hendel (1998, p. 5) emphasises the role of technological development in book design:

What once took hours and days can now be done in minutes ... I used to spend hours, even days, drawing out a title page. I was consummately proud of my skill at rendering letters. Now the computer has made all of this unnecessary. No longer dependent on my ability to draw letters, I can see precisely what the letters look like and can control every detail of their final placement.

Most importantly, technological development has brought about a number of tools to assist the production. To refer back to the model of change shown in Figure 3.1, the available tools and technology also shape the structure of a multimodal artefact, as they affect the process of creating the artefact. For example, the cutting and pasting of text and images on millimetre paper has been replaced by desktop publishing (DTP). In DTP, the same process happens without the need for a material substrate, as the design process takes place in a digital environment and the end product of the process can be sent directly to the press. What this opens is the possibility of increased experimentation in design, the impact of which I shall describe below.

However, a discussion of the current production methods and tools should be preceded by a look into the past. Previously, graphic design did not only require the skill of visual composition, but also the art and craft of making things by hand. This process involved the careful positioning text and images on millimetre paper to prepare a plate, which could be then used to produce a print run in the press. Let us now look at two specific examples selected from the annotated corpus to illustrate the design process in the past.

The first example, shown in Figure 6.5a, is from a series named *Helsinki’s Four Tourist Islands*, first published in 1967. The example shows a part of a map, which helps the reader to travel to the “tourist islands”. In this case, the point of interest is how the map and the accompanying graphic elements are put together. The map, upon which the graphic elements are imposed, is a nautical map intended for navigation! On the map, the transparent blue lines indicate the ferry routes. The



(a) Map in *HFT* 1967

(b) Text in *SMH* 1982

Figure 6.5: Two examples of technological development

destinations, in turn, are marked by the round, coloured symbols. These symbols also reappear close to the destination descriptions elsewhere in the brochure — I will explain their function shortly below.

Given the tools available in 1967, the design solution in Figure 6.5a is simple yet effective. At the same time, it is important to understand that each element — the map, the blue lines, the symbols — had to be separately manufactured and put together in the final design. Moreover, any experimentation with different designs would have warranted the same process. For this reason, graphic design remained a very concrete process until the invention of DTP.

To illustrate the changes resulting from the invention of DTP, we can trace the realisation of the same map in the *Helsinki's Four Tourist Islands* series in the examples shown in Figure 6.6. As Figure 6.6a shows, the same technique from 1967 was still used in 1972 (cf. Figure 6.5a). Fast-forward twelve years; the map in Figure 6.6b has undergone significant changes, plausibly due to the introduction of DTP. Firstly, the nautical map has been replaced with a map of the coastline and only major neighbourhoods are named on the map. Considering the map cannot be used for orientation — and certainly not for navigation — it was likely created for the brochure in question.

Four years later in 1988, the realisation of the map has changed again (see Figure 6.6c). In particular, the destination names are now accompanied by coloured boxes, which also appear elsewhere close to the header of the respective sections of the brochure. This is a kind of implicit navigation structure designed to help in the use of the brochure (Hiippala 2012b, p. 1505). In addition, the shape of both the end point and the line have changed. Unfortunately, the series was apparently discontinued after the year 1988, so further examples could not be found. In any

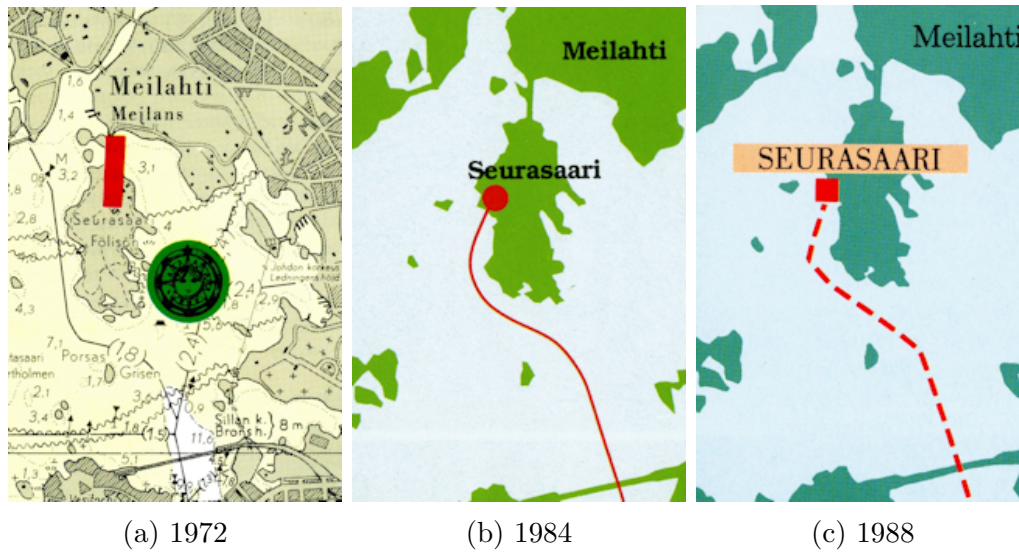


Figure 6.6: Maps in *Helsinki's Four Tourist Islands*

case, the introduction of DTP appeared to foster experimentation with designs and visual appearance at least in this series.

Let us now return to the other example given in Figure 6.5b. This brochure, *Sculptures and Monuments in Helsinki* (1982), is an interesting example in the annotated corpus: it is the only brochure that was not typeset. Instead, the brochure was written with a typewriter. As a result, the variation in typography is limited to small and capital letters, and possibly to underlined text. The chosen production method is therefore very limited in terms of typography, especially if contrasted with today's possibilities. What I wish to emphasise here is how far the design of multimodal artefacts has advanced: many of the technical constraints in production have largely disappeared.

However, the major implication of the discussion above is the following: a multimodal analyst is only provided with the end product of the design process. The analyst cannot see the various iterations of the design process; nor the analyst cannot know if the end product was created by multiple authors (cf. E. et al. 2011). Today, a copywriter and an art director are likely to share the work — not to mention other supporting roles that may be involved in the process, such as project managers, editors and DTP operators. In future, it is possible that joint work with ethnographic methods may be used to investigate the design processes as they unfold (see e.g. Kress 2011). At the time, what can be said with relative certainty is that the principles of the design process have changed.

In conclusion, it may be suggested that the need for a material substrate to realise a design encouraged a careful and considerate approach to the process



of graphic design in the past. In contrast, the possibility of instant feedback and redesign — enabled by DTP — may have fundamentally changed the craft. To the multimodal analyst, the important question is whether this possibility to experiment and redesign has led to changes in the brochures. It should also be noted that reworking the visual appearance of a brochure may also involve altering its multimodal structure in terms of the layout structure and the rhetorical structure: I shall follow up on this issue in Section 8.5.

## 6.4 Summary

In this chapter, I have laid out the groundwork for the further analysis of the tourist brochure as a multimodal artefact. Addressing the characteristics of the brochure as a medium is essential in preparation for the analysis of multimodal structure in the subsequent chapters. The capabilities and characteristics of the selected medium — the brochure — are necessary for distinguishing and identifying clearly the various contributions to the multimodal structure of the brochures. I will now summarise the three issues discussed in this chapter.

To begin with, Section 6.1 explored the characteristics of the brochure as a medium, which carries the genre of the tourist brochure. After establishing the necessary theoretical prerequisites, I studied certain features of the medium, such as the method of binding and the fold geometry, in order to establish how the brochures exploit the material substrate of a printed page to create space for their content. The analysis established that staple-bound brochures have a higher page count (mean = 24.6), whereas in the folded leaflets without staples the number of pages is lower (mean = 6.8). I concluded that these general observations may have implications for the subsequent multimodal analysis, if differences are found between the structures of brochures and leaflets (see Section 8.5).

In Section 6.2, I addressed the issue of advertising in the tourist brochures. Although the tourist brochures themselves constitute a form of advertising, a number of brochures ( $n = 6$ ) also contained classified advertisements. This development appeared only after the year 2000 in staple-bound brochures with a high number of content pages. Furthermore, the classified advertisements also expanded the functionality of the tourist brochure as an artefact, because the brochure could be used to redeem discounts at the locations promoted in the classified advertisements. To conclude, I also argued that advertising should be considered a medium-based phenomenon in the process of modelling the structure of a multimodal artefact. This means that the analysis of the advertisements' structure should not be incorporated into that of the tourist brochures as a genre, because the genre and the advertisements do different kinds of communicative work.

The concluding Section 6.3 considered two aspects of production related to the tourist brochures. I first questioned the much-discussed ‘visual turn’ in communication and evaluated the proposal by drawing on the annotated corpus: for each double-page, I calculated the number of visual base units and the amount of layout space they occupy. The analysis showed the tourist brochures have been fairly visual between 1967 and 2008, with the average percentage of layout space dedicated to graphic elements exceeding over 50%. It appears that the medium of a brochure affords a high degree of visuality, but the annotated corpus showed that the pages take many forms: some are purely linguistic, some are mainly visual, and some are in between. In addition, I discussed and exemplified the impact of technological development on the tourist brochures. In preparation for the structural analysis, I asked whether desktop publishing has fostered experimentation that resulted in visible changes in the multimodal structure of the brochures. This observation added another point of focus for the subsequent analysis.

With the three points described above in mind, the following chapter continues with an in-depth analysis of the brochures and their structure on the basis of the annotated corpus.

# Chapter 7

## The content and its structure

This chapter provides a detailed analysis of the content and its structure on the basis of the annotated corpus. For this purpose, I use the notion of a semiotic mode to show where the decisive choices that affect the structure of a multimodal artefact are made.

I have already established that language and image play an important role in constructing a mental image of the destination for the tourist (Molina and Esteban 2006). In more concrete terms, the brochures typically introduce carefully selected aspects of the destination to the reader. Valdeón (2009), for instance, identifies several common topics in the tourist brochures, such as history, gastronomy, geography and climate. In his analysis, Valdeón explores the metafunctional characteristics of the linguistic descriptions associated with these topics. However, in relation to the multimodal characteristics of the tourist brochures, Valdeón (2009, p. 26) observes that:

The layout combines visual and linguistic inputs, thus creating poly-semiotic texts. However, the distribution of information in the brochures and leaflets differs considerably.

Although the metafunctions were not considered to be a viable point of departure for modelling the structure of a multimodal artefact (see Section 2.3.3), this dissertation is very much concerned with the distribution and structuring of the content (“information”) in the tourist brochures. For this reason, the semiotic modes used to realise the content are subjected to a careful examination in this chapter. At this point, it is also useful to recall the model of the tourist brochure as a genre and its communicative goals (see Section 4.3.1).

In short, I want to find out whether specific semiotic modes are geared towards fulfilling distinct communicative goals as a part of a broad rhetorical strategy (cf. Bateman forthcoming, pp. 8-9). If this is indeed the case, I want to find out how exactly the semiotic modes accomplish this task. By carefully taking apart

the semiotic modes in this chapter, the dissertation stands in good position for studying the interpretation of the brochures in Chapter 8. I shall begin this task with a discussion of text-flow: a foundational, language-based semiotic mode.

## 7.1 Representing the destination using text-flow

As Section 3.4.1 established, the semiotic mode of *text-flow* deploys the full meaning potential of language and can thus be used in various communicative contexts. In this section, I will focus on the structure and functions of text-flow in the tourist brochures. More precisely, I aim to concentrate on the following issues: (1) what kind of content is communicated using text-flow, and (2) how is this content structured? For answering these questions, the main source of data is the rhetorical layer of the GeM-annotated corpus, which is complemented by data from the layout layer to describe the hierarchical structure of the artefact. This enables the observation of the rhetorical structure in its context of occurrence, that is, in a paragraph or another kind of layout unit.

I also wish to underline that unlike the previous studies of tourist brochures (see Section 1.4.1), I focus mainly on the broad, textual aspects of language and how it supports the communicative goals of the brochures. In this way, the application of RST as a part of this dissertation is also a novel contribution to linguistics, because previous research has not explored the rhetorical structure of the tourist brochures. My analysis, of course, is conducted with multimodality in mind.

To exemplify why RST-based linguistic and multimodal analysis is necessary for the task at hand, I begin with a short keyword-in-context list retrieved from the annotated corpus, where each line includes an instance of ‘Helsinki’:

```
Helsinki was founded in 1550 by Gustav Vasa ... (1969-hdb/u-1.25)
Helsinki is not far away ..... (1972-hdb/u-2.03)
This is the oldest building in Helsinki ..... (1978-wtz/u-1.08)
The clock tower of Helsinki Railway Station ... (1980-tgh/u-4.33)
Helsinki keeps going all year round ..... (1998-weh/u-3.34)
```

A traditional, monomodal approach to corpus linguistics would most likely focus on some grammatical aspect of Helsinki in each clause (see e.g. Guijarro and Hernandez 2001; Bonelli and Manca 2002). In contrast, a model of artefact structure has to identify the functions of these clauses in relation to the other elements on the page. The clauses above may be used to exemplify this issue: they perform the functions of titles, captions or form parts of the linguistic content. When the analysis expands to cover the entire multimodal page, it soon becomes evident that the discursive functions of text segments cannot be studied effectively using simple

keyword-in-context lists. This does not, however, rule out using the keyword-in-context lists for detailed linguistic analyses of the content in multimodal research, as Francesconi (2011) has done.

Table 7.1: Rhetorical relations in the annotated corpus

JOINT	656	CONDITION	5
ELABORATION	447	MEANS	5
ENABLEMENT	312	CONCESSION	5
RESTATEMENT	121	PURPOSE	4
LIST	55	VOLITIONAL-RESULT	4
IDENTIFICATION	51	CIRCUMSTANCE	3
BACKGROUND	35	NONVOLITIONAL-CAUSE	3
SEQUENCE	30	ANTITHESIS	3
EVALUATION	30	INTERPRETATION	2
PROPERTY-ASCRPTION	18	NONVOLITIONAL-RESULT	2
MOTIVATION	15	SOLUTIONHOOD	1
EVIDENCE	14	JUSTIFICATION	1
PREPARATION	9	CONTRAST	1
LOCATION	6	SUMMARY	1

In order to proceed, I shall now look at the rhetorical relations ( $n = 1839$ ) in the annotated corpus (see Table 7.1). At this point, however, a word of warning is appropriate. The RST annotation was done by one annotator, without any measures to validate the annotation by using other annotators and comparing the results (for a discussion of the analysts and the analytical process, see Taboada and Mann 2006*b*, pp. 443-445). For this reason, I will not describe the multimodal structure of the tourist brochures on the basis of the relation counts, but rather by observing the distribution of each relation, while paying particular attention to their contexts of occurrence.

Interestingly, the most frequent relation to occur in the corpus is JOINT, whose definition states that no rhetorical relationship holds between the participating segments. JOINT is followed by ELABORATION and ENABLEMENT; together the three relations amount to 77% of the rhetorical relations in the corpus. However, the broad overview of the relations and their distribution in Table 7.1 does not reveal their functions and contribution to the discourse structure. Moreover, Table 7.1 does not show how different relations are distributed in the artefacts that make up the annotated corpus. Therefore, in the following discussion, I attempt to capture the structure of text-flow, while simultaneously laying out its functions.

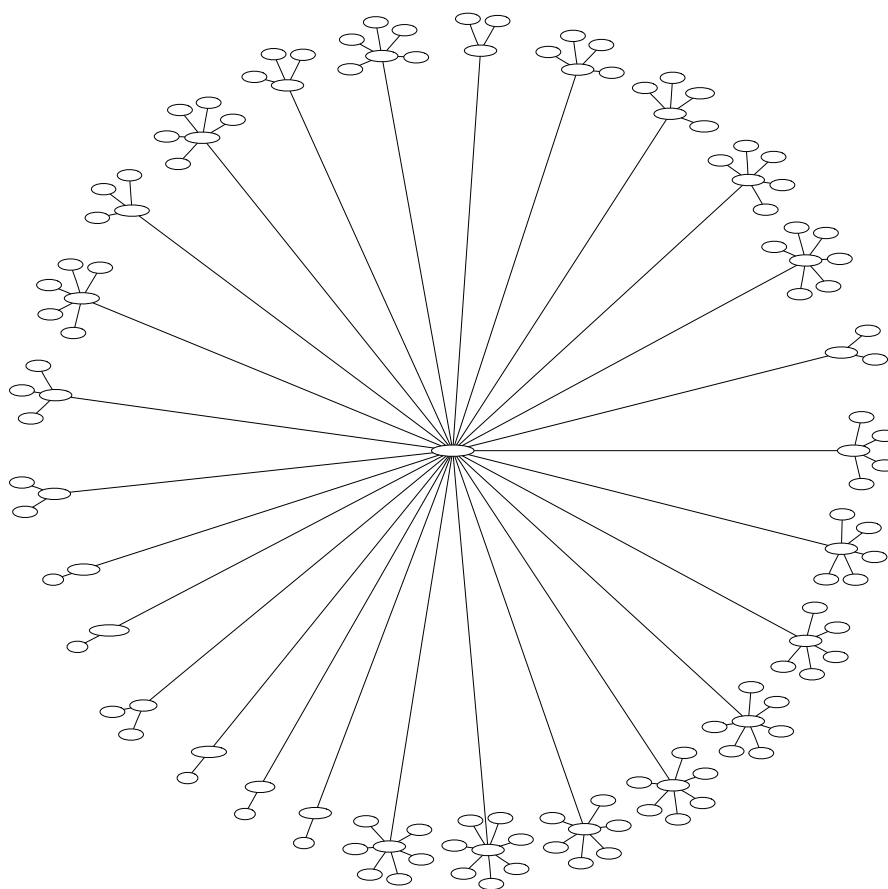


Figure 7.1: The layout structure in *MEH* 1967

### 7.1.1 Joint: the use of tables

I shall begin with the multinuclear relation of *JOINT*, which is the most frequent relation in the annotated corpus ( $n = 656$ ). Concerning the occurrence of *JOINT*, it should be noted that three brochures, in which *JOINT* is the most common relation, contain 71% of the *JOINT* relations in the annotated corpus.<sup>1</sup> These brochures are characterised by the fact that they (1) communicate to a very specific audience, such as the organisers of meetings and conferences, (2) contain a wealth of information with an average of 169 base units per content page, and most importantly, (3) structure the information using tables, which typically condense meanings (Baldry and Thibault 2005, p. 64).

These brochures do not argue for their content rhetorically. For this reason, a table is an effective way of communicating the content, because the tables structure the content logically by placing it in columns and rows in the layout. This can be

---

<sup>1</sup>*CEH* 1967; *MEH* 1967; *SMH* 1982.

seen in Figure 7.1, which shows the layout structure of the brochure *MEH* 1967. This layout structure resembles a ‘hub-and-spoke’ structure. If contrasted with the structure shown earlier in Figure 3.2, it may be observed that the layout structure in Figure 7.1 lacks structural depth. This sets certain limitations to the use of tables: I will now make these limitations explicit below.

The visualisation in Figure 7.1, retrieved from the annotated corpus using XQuery and transformed into DOT, shows how the brochure uses a hub-and-spoke structure to organise the information. Most of the nodes connected to the parent node, which is located in the centre, include the same child nodes which carry the genre content: name of the event, date, place and how to acquire more information. Visually, the wealth of information contained in this table is made accessible by its organisation into columns and rows (Lemke 1998, p. 96). Yet the layout structure remains shallow: elaborating the content presented in the table would require introducing additional child nodes below the current child nodes.

Because the grammar of the language used in tables is stripped to a bare minimum, tables cannot argue for the content rhetorically. This also explains the high number of JOINT relations: a reduction in the stratum of lexicogrammar is also bound to affect the more abstract stratum of discourse semantics (cf. Martin 1997, p. 6), which the application of GeM RST seeks to capture. After all, it is the lexicogrammar that provides language with its immense meaning potential, which enables a rhetorical structure to take form. Moreover, the lack of argumentative structure may also result from the intended audience of the brochures. As the brochures are directed to specialist audiences, they are more oriented towards giving practical information and therefore aim less at ‘selling’ the destination to the reader.

To conclude, the rhetorical layer of the GeM model is not the most effective tool for analysing the structure of multimodal artefacts, in which the layout structure follows a hub-and-spoke pattern. For the most part, these artefacts consist largely of tables or lists. The layout layer of the GeM model, however, provides the means to study the structure of what may hereby be termed *table- and list-driven artefacts* in considerable depth. For example, the realisation information could be used to highlight the position of access structures — such as headers and other segments with typographic emphasis — in the nodes of the layout structure (cf. Waller 1987). However, owing to limited space, this falls outside the scope of this dissertation. Nevertheless, the usefulness of rhetorical analysis will become evident when the analysis moves towards other types of text-flow, for example, in the connection with the relation of ELABORATION, which I will discuss next.

## 7.1.2 Elaboration in descriptive texts

As Table 7.1 shows, ELABORATION accounts for 24% of the total relations in the corpus ( $n = 447$ ). ELABORATION is also the most common relation in 17 out of 30 brochures. Consequently, we may ask why the relation of ELABORATION is so frequent in the tourist brochures? At this point, it is necessary to recall that Stede (2008, p. 318) has criticised ELABORATION as a particularly vague and imprecise relation. He writes:

[M]any annotators seem to resort to ELABORATION as a “default”, so that the presence of this relation in some RST tree can result either from the spans being in a genuine, “good” ELABORATION relation (and the annotator confidently assigned it) or from a perceived unclear relationship between the spans, which is somehow also covered by ELABORATION.<sup>2</sup>

I would argue that any judgement regarding ELABORATION — or any other RST relation — needs to be sufficiently informed by the context where the relation is considered to hold. After all, this is a key principle of RST as a theory of text organisation. If there is a risk of using ELABORATION more often than necessary, a case-by-case study of the relation should be able to detect the behaviour of “defaulting” to ELABORATION in problematic analyses.

As I will show below, the relation of ELABORATION appears to have several functions in the data. ELABORATION does not only add information to a previously introduced topic by establishing a relation between two RST segments: the relation also occurs between a segment and a *span*. By incorporating another span in the RST structure, ELABORATION facilitates the introduction of rhetorical relations to the structure. Let us now look at an example below.

Figure 7.2 shows the rhetorical structure of an instance of text-flow in *WEH* 1998, which is made up of several paragraphs. This is reflected in the layout structure, as the text consists of a layout chunk with multiple child nodes: the header (**s-3.24**) and three paragraphs (**s-3.25**, **s-3.26-9** and **s-3.30-2**). In terms of the rhetorical structure, I would suggest that this text is a coherent entity with a specific function that describes a common topic in the tourist brochures: the climate (Valdeón 2009, p. 26). Yet to establish how such coherence is achieved, the rhetorical structure of the text needs to be subjected to a closer examination.

To begin with, the paragraph in the segment **s-3.25** is central to the text. All other paragraphs elaborate this paragraph, as indicated by the topmost ELABORATION span in Figure 7.2. This topmost span embeds three other spans: an ANTITHESIS span and two ELABORATION spans. Together, the ANTITHESIS and

---

<sup>2</sup>Note that I have applied the typographic conventions used in this dissertation to this quote.



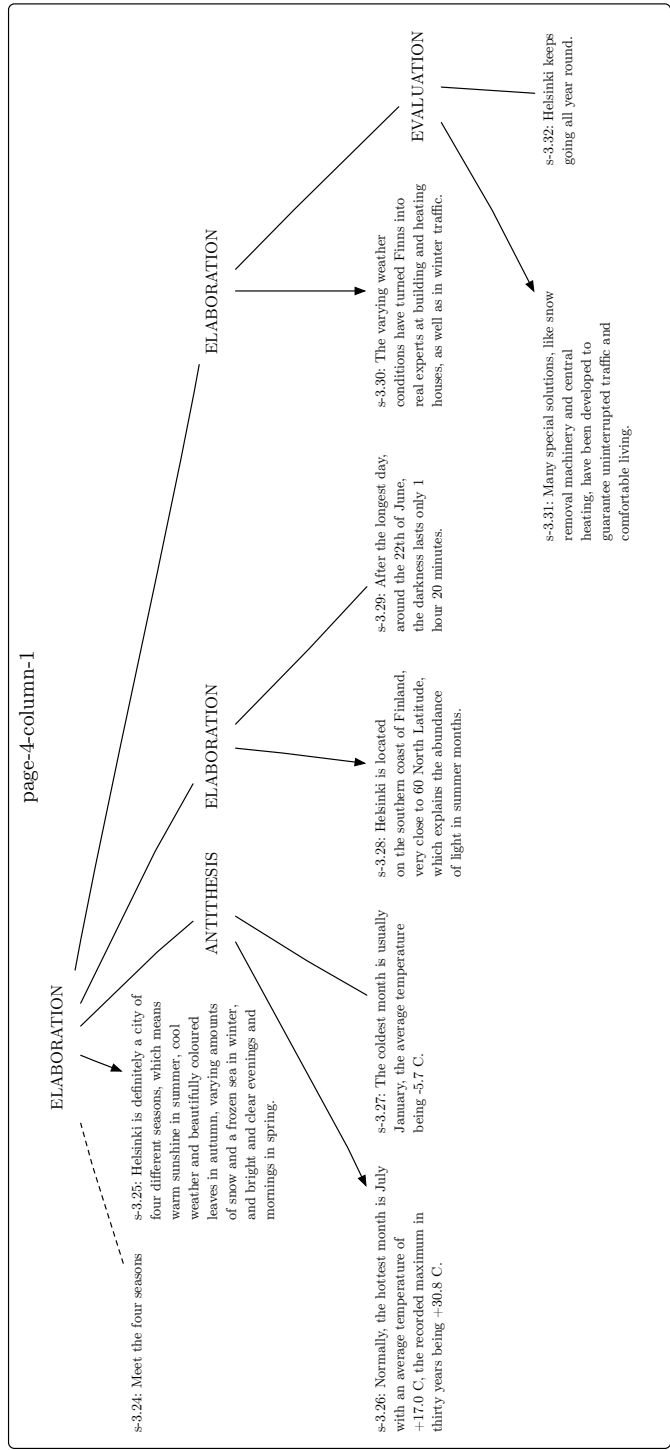


Figure 7.2: Examples of ELABORATION in *WEH* 1998

the first ELABORATION span constitute a paragraph. What these embedded spans do is they endow the description with a certain ‘depth’, which allows the introduction of additional topics related to climate. This is further illustrated by the third ELABORATION span, which embeds an EVALUATION span that explains how the citizens of Helsinki have learned to live with the local climate.

One possible explanation for the rhetorical structure in Figure 7.2 arises from the evaluative function of the tourist brochures. Considering that the brochures usually aim to provide a positive appraisal of the destination (Francesconi 2011, p. 344), they require a mechanism for introducing and highlighting selected aspects of the destination. From the perspective of discourse structure, the paragraphs may be considered macrostructures with specific functions that provide the required mechanism.

In this case, the paragraphs introduce (1) the general topic and describe (2) the climate, (3) the location (4) and their combined effect on the city and its inhabitants. These paragraphs may be further split into microstructures that consist of RST segments and their interrelations (Longacre 1992, pp. 112-114). From this position, the investigation could obviously be extended all the way down to the stratum of lexicogrammar (cf. Hiippala 2007). Together, the paragraphs and the segments that constitute them illustrate the meaning potential and flexibility of text-flow in this aspect: the same structural pattern could be used to introduce a completely different topic.

Consequently, I set out to investigate how often another span acts as a satellite in an ELABORATION relation by querying the annotated corpus for the instances that match the aforementioned criteria. The result showed that in 81 out of the 447 ELABORATION spans (18%), the satellite consisted solely of another span. In 197 ELABORATION spans (44%), the satellite consisted of a number of RST segments and another span. I would thus suggest that the purpose of ELABORATION is to do more than simply add detail to the description. While also adding detail to the topics introduced in the text, the relation simultaneously enables the introduction of other rhetorical relations by embedding additional RST spans.

It may be suggested that this kind of recursive *deep elaboration* is a particular rhetorical configuration to enrich and liven up the descriptions in the tourist brochures. Finally, if the recursive use of ELABORATION were to be studied further, the application of classical RST could be beneficial (see Section 3.2.3.1). Determining the relations between clauses — not sentences — would provide the appropriate detail for focusing exclusively on the structure of linguistic discourse. Keeping in mind the multimodal goals of this dissertation, I will now move on to discuss the next relation: ENABLEMENT.

### 7.1.3 Enablement: supporting the descriptions

Table 7.1 indicates that ENABLEMENT is the third most frequent rhetorical relation with 312 instances (17%) in the annotated corpus. As a rhetorical relation, ENABLEMENT has an important function in the tourist brochures. ENABLEMENT spans provide the reader with information about typical touristic activities: access to the destination, social and cultural activities, tips, itineraries, what to do in the case of an emergency, and so on (Thurlow and Jaworski 2010, p. 193). In relation to the descriptions discussed in the previous section, ENABLEMENT has an important *supporting* role, making it possible for the tourist to pursue the activities associated with and proposed in the descriptions. This is exemplified in Figure 7.3, which shows an extract from an instance of text-flow in *HFT* 1972.

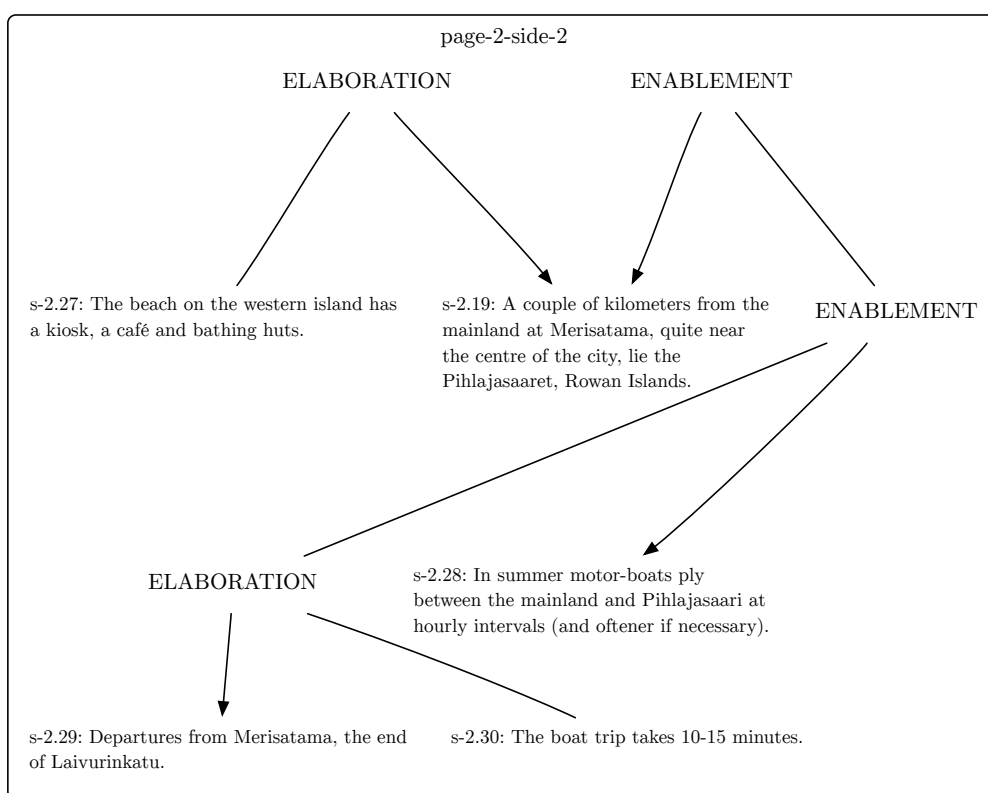


Figure 7.3: Enabling the tourist to perform activities in *HFT* 1972

The main point of interest in Figure 7.3 is the ENABLEMENT span, with the segment **s-2.19** as the nucleus. The satellite embeds two more spans, ENABLEMENT and ELABORATION, which are used to provide detailed instructions on how to reach the destination. In short, this rhetorical configuration tells the reader how to reach the islands by boat, from where the boats depart, and how long

the journey takes: all the essential information that the tourist needs. Extensive descriptions are not always provided — in many cases, the ENABLEMENT spans provide the bare minimum, such as a phone number with street and website addresses (see Figure 7.4).

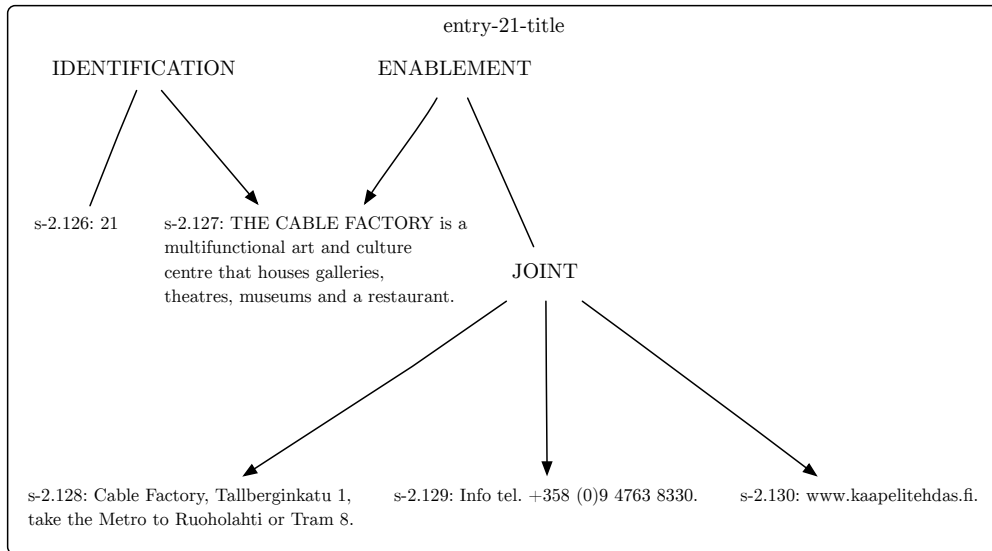


Figure 7.4: Enabling the tourist to perform activities in *BNO* 2006

In conclusion, I wish to restate that the role of ENABLEMENT, particularly in relation to the the location and destination descriptions, is largely supportive. The ENABLEMENT spans provide practical information, while the persuasive work mainly is done by the descriptions. For the actual guidance around the destination, the responsibility lies mainly with the multinuclear relation of SEQUENCE, which I shall describe next.

#### 7.1.4 Sequences: history and guidance

Here I want to pay attention to the special role of SEQUENCE as a rhetorical relation in the tourist brochures, although only 30 instances of SEQUENCE appear in the corpus (see Table 7.1). It appears that the relation is multifunctional and primarily used for two purposes: (1) to provide an account of historical events, and (2) to guide the reader through a sequence of activities, for example, while visiting a series of locations within a destination. To generalise, both functions involve events which unfold in time.

Figure 7.5 shows the first example, an extract of a *historical* SEQUENCE in *HFT* 1967, which provides a brief history of the 18<sup>th</sup> century maritime fortress of Suomenlinna. The multinuclear relation of SEQUENCE is used to provide the

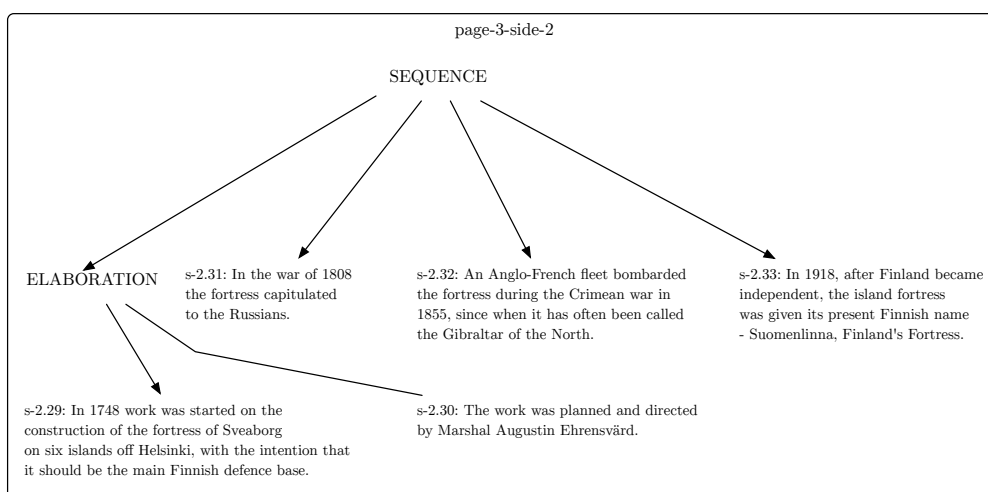


Figure 7.5: Telling the history of Suomenlinna using a SEQUENCE in *HFT* 1967

reader with a list of important events in the history of the described location (s-2.29–33). ELABORATION, in turn, provides additional information on Ehrensvärd, a Swedish field marshal involved in the planning and construction of the fortress. Such use of a historical SEQUENCE was found in a total of nine brochures in the annotated corpus.

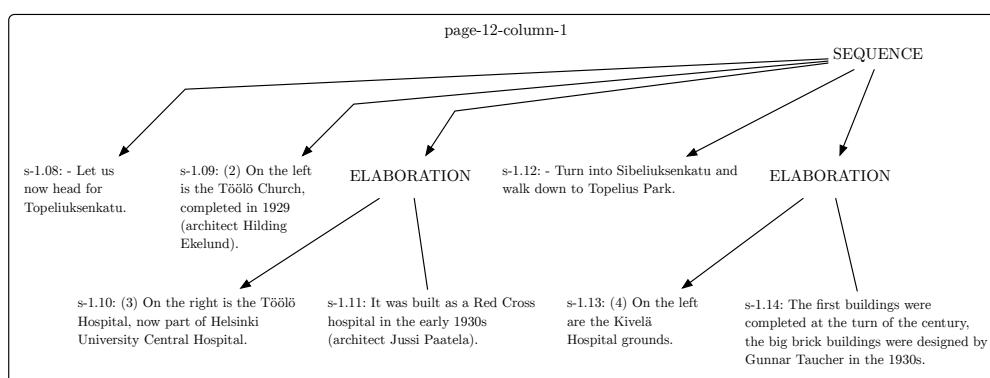


Figure 7.6: Guiding the reader using a SEQUENCE in *SHF* 1987

The second example of SEQUENCE illustrates how the reader is guided through the destination using a “stop-look-see” discourse strategy (Enkvist 1991, p. 9). Figure 7.6 shows an extract of a *guiding* SEQUENCE, which instructs the reader through the district of Töölö (s-1.08–14). This is a typical step-by-step procedure with ELABORATION providing additional detail. This rhetorical configuration may be more typical of a guide book than a tourist brochure, considering its relatively low count in comparison to other relations in the corpus (for guide books, see

Waller 1984; Enkvist 1991). A SEQUENCE used for guidance was found in a total of five brochures. Additionally, one brochure included both historical and guiding use of a SEQUENCE.

It should be noted, however, that the brochures with a high number of SEQUENCES (8 in *WTZ* 1978; 6 in *WTH* 1983) are specifically designed for guiding the tourist around the destination (both are a part of a series entitled *A Walking Tour in Helsinki*). In these brochures, a top-level SEQUENCE span embeds multiple SEQUENCE relations. These ‘sub’-SEQUENCES are then elaborated to a considerable degree of detail using the rhetorical strategy of deep elaboration, which I described above in Section 7.1.2. Structurally, the rhetorical configuration of these brochures resembles the SEQUENCE in Figure 7.6, although the elaboration of detail is far more expansive.

This kind of top-level SEQUENCE is also reminiscent of the organisation of procedures described in André and Rist (1995), where the topmost span is responsible for keeping the overall procedure intact, while the actual processes take place in the embedded structure (for another example, see Matthiessen 2007, p. 34). However, in contrast to the procedures in André and Rist (1995), these brochures are exclusively verbal apart from certain graphic elements on the cover page and the accompanying map. This, in turn, raises an important point: language alone is not sufficient for guiding the tourist. Instead, moving around the destination requires a map — a visual representation, which acts as a “graphical user interface” to the city (Vertesi 2008, p. 25), which I will discuss in Section 7.3.3. As for now, I will move on to discuss the use of graphic elements alongside text-flow in the tourist brochures.

## 7.2 Graphic elements in the tourist brochures

I will now continue with the description of the content and its structure in the tourist brochures by extending the analysis to photographs, illustrations, maps and other graphic elements in the data. Specifically, my aim is to find out which visual semiotic modes are used to realise the genre of a tourist brochure. At this point, however, it is necessary to briefly return to the semiotic mode of *image-flow*. I have already discussed image-flow in Section 3.4.2, in which I emphasised its role in organising images into meaningful sequences. The annotated corpus, however, did not contain any justifiable instances of image-flow. Therefore, the main focus will be on the graphic elements and their interaction with *text-flow*, and eventually on their combinations as a part of *page-flow*. The discussion starts with an overview of the graphic elements.

Initially, it is important to establish what kinds of graphic elements are included in the data. As Section 6.3.1 established, the medium of a brochure has been

Table 7.2: The distribution of graphic elements in the data

	Number	Percentage
Photographs	158	44.4%
2D elements	109	30.6%
Illustrations	61	17.1%
Maps	28	7.9%
Total	356	100%

fairly visual throughout the period studied in this dissertation (see Figure 6.4). Additionally, the examples covered in the previous chapters have included both photographs and illustrations in various forms and sizes. However, to achieve a view of their distribution in the data, the annotated corpus needs to be consulted. This is the first step in considering the function of graphic elements as a part of the multimodal structure.

Table 7.2 shows the distribution of graphic elements in the annotated corpus: a query of the realisation information returned a total of 356 graphic elements. The most frequent graphic element type is a photograph, which accounts for 44.4% of the identified graphic elements. Following the photographs, the most common graphic element types are two-dimensional elements (30.6%), illustrations (17.1%) and maps (7.9%). As the use of the graphic elements remains consistent during the period represented in the data, I shall now proceed to discuss how these graphic elements function as parts of the brochures' multimodal structure. The first graphic element to be discussed is the photograph.

### 7.2.1 Photographs in image-text-complexes

Previous research on multimodality has made several observations about the content of photographs in tourism discourse. Machin (2004, p. 330) points out that unlike the highly decontextualised photography in contemporary image banks, the tourism-related photographs still retain contextual meanings, because they can be placed in a geographical context. Traditionally, the geographical context has been invoked by portraying the established landmarks and tourist attractions. The work of Jokela on the visual representation of Helsinki between the years 1954 and 1963 lends support to this claim, as “one-third of the analysed images depict the old monumental center” (2011, p. 59).

In Hiippala (2007), I proposed that photography in the tourist brochures may also be shifting towards representing the values associated with certain ‘lifestyles’, as opposed to only portraying landmarks and touristic attractions. Such a shift

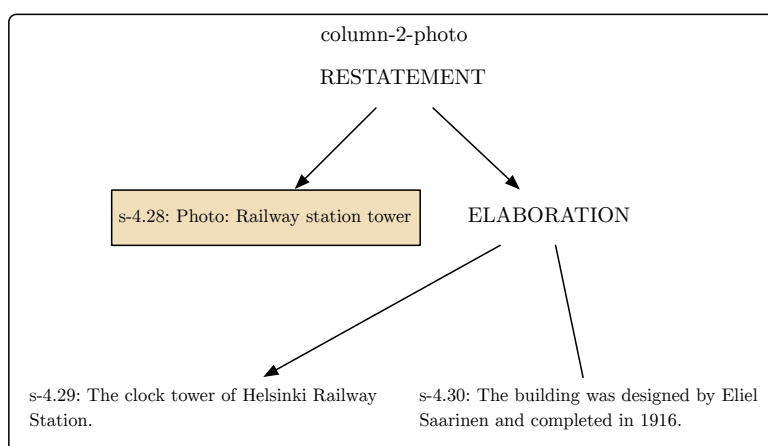


Figure 7.7: An image-text complex with a RESTATEMENT relation in *TGH* 1980

would be consistent with the discourses of destination branding, which seek to associate the destination with a set of values (Francesconi 2011, p. 342). Yet these values are not easily communicated in the visual semiotic modes by portraying concrete entities, which has led to the introduction of ‘conceptual’ photographs to the tourist brochures (see Section 3.3.3).

Although I do not explicitly analyse the content of the photographs, the possible shift in the use of the photographs should be kept in mind when studying the structure of the tourist brochures. The possible shift warrants particular attention on how the photographs are incorporated into the overall multimodal structure. In order to proceed, I shall now continue the discussion from the perspective of structure, beginning with the photographs that are accompanied by captions.

Following Kvåle (2010), the concept of an image-text-complex shall act as the point of departure for the analysis of photographs and text-flow (see Section 3.5.3). The reason to consider the image-text-complex is the high semiotic potential of its structure. What I mean by this potential may be exemplified by Figure 7.7, which shows an image-text-complex with a RESTATEMENT relation with two nuclei: a photograph (s-4.28) and a caption. The caption consists of two RST segments (s-4.29-30) that are joined by the rhetorical relation of ELABORATION.

In this case, the discourse semantics of the image-text-complex are seemingly straightforward: the close proximity of the text and image signals that they should be interpreted together (see also the example in Figure 3.11 in Section 3.5.3). This interpretation is further reinforced by the hierarchical structure, as the entire image-text complex is formed by the child nodes of the same parent node in the layout structure, as indicated by the bounding box around Figure 7.7.

The structural simplicity of the image-text-complex may be deceiving, because its meaning potential lies in the discourse semantics of the image-text-complex.



Recall that the discourse semantics also guide the interpretation of the entire pages, which follow the logic determined by the active semiotic mode (see Section 8.4.1). Structurally, the image-text-complex is often just one part of the page and its discourse semantics provide the key to interpreting the *structure* of the image-text-complex.

It is possible, however, to insert different semiotic resources into the structure provided by the image-text-complex, which opens up a wealth of semiotic possibilities. This structure is precisely where the semiotic potential of the participating semiotic modes can manifest itself in phenomena such as multimodal metaphor (Forceville 1996; O’Halloran 1999a; Van Mulken et al. 2010), cohesion and intersemiosis (Royce 1998, 2007; O’Halloran 2008b; Liu and O’Halloran 2009). To exemplify, the work of Caple (2009a) on “image nuclear news stories” shows how complex intersemiotic meanings are embedded into a simple multimodal structure — a structure that closely resembles an image-text-complex. A further example can be found in Knox’s (2007) study of “newsbites” on newspaper websites.

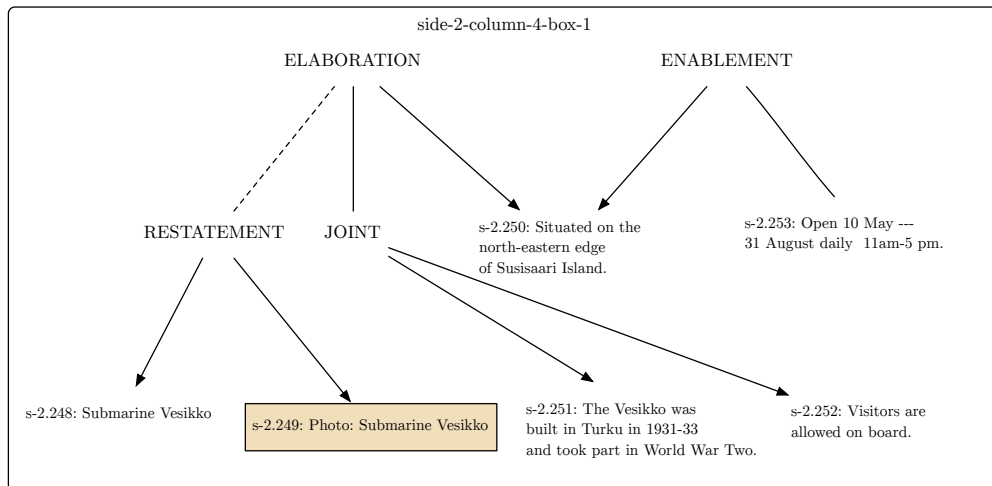


Figure 7.8: An image-text complex in *SSH* 1986

The widespread use of the image-text-complex also warrants attention to its use in the annotated corpus. Figure 7.8 shows another type of an image-text-complex in the tourist brochures, which I hereby term an *illustrated description*. As the bounding box indicates, all the rhetorical segments exist under the same node in the layout hierarchy. In terms of structure and its discourse semantic interpretation, the captioned photograph in Figure 7.7 and the illustrated description in Figure 7.8 both rely on spatial proximity in the layout to signal that they are rhetorically connected.

However, the rhetorical configuration of the illustrated description in Figure 7.8 is different than that of the photograph and its caption in Figure 7.7. A

RESTATEMENT relation holds between the header (**s-2.248**) and the photograph (**s-2.249**), which together act as a TITLE span for the description. This description, in turn, consists of the nucleus (**s-2.250**) and its satellites (**s-2.251-3**), which provide additional detail and information on how to access the described location. Unlike the captioned photograph in Figure 7.7, the illustrated description does not explicitly establish a relation between the photograph and a specific RST segment in the description. In many cases, the proximity and presence of a header and an image are considered sufficient to establish a rhetorical relation between the participating segments. It may be argued that this is supported by the staging of the tourist brochure as a genre: together, the header and the photograph fulfil the stage of identification (see Section 4.3.1). This interpretation is also reinforced by the discourse semantic structure of the image-text-complex.

Why, then, the tourist brochures do not always choose to emphasise and explicitly signal the rhetorical structure, particularly in the case of image-text relations? In the search for an answer, it is important to keep in mind the economy of *space* on the page. In the tourist brochures, the semiotic space is a valuable commodity, because the brochures need to be small-sized for easy distribution. As I have pointed out in Hiippala (2013), changes in the rhetorical structure also affect the layout and vice versa, for example, in deciding whether to provide captions to accompany the photographs. The captions take up layout space, and in some cases, the efficient use of layout space may override the need for signalling rhetorical relations. This sets the stage for further discussion, which focuses on conceptual photography, that is, generic photographs without captions.

## 7.2.2 Rhetorically weak: conceptual photographs

As I observed above, not all photographs in the data are accompanied by captions. In fact, many of the photographs stand alone in the brochure structure, without an explicit rhetorical relation to any specific segment of text or another image. I refer to these photographs as conceptual, suggesting that their function is to provide the reader with a general view of the portrayed destination or location and the abstract values and meanings associated with the target of description. Below, I will show how these conceptual photographs are integrated into the multimodal structure of the tourist brochures.

Figure 7.9 shows a configuration which integrates multiple conceptual photographs into the multimodal structure of the brochure using the header (**s-4.04**). Like in the illustrated descriptions described in Section 7.2.1, the header is the only rhetorical ‘point of contact’ between the photographs and the accompanying text-flow, which provides a detailed description of the location (not shown, continues with the dashed TITLE line). This description does not, however, establish an explicit connection with the photographs. Not surprisingly, several cohesive ties

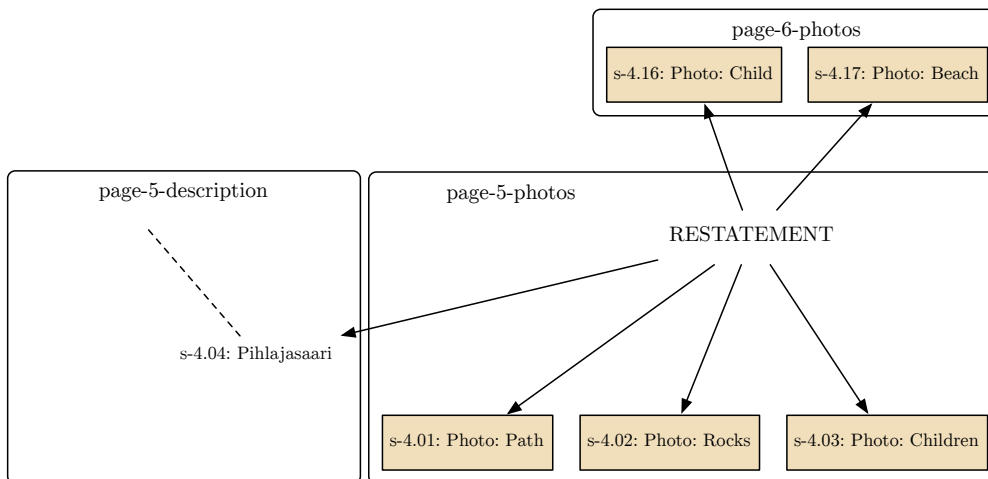


Figure 7.9: Conceptual photographs in *HFT* 1984

may be nevertheless identified between the text and the photographs, exemplified by nominal groups such as *sandy beaches* (cf. **s-4.17**) and *beautiful footpaths* (cf. **s-4.01**).

However, I do not consider these cohesive ties strong enough to justify a claim that a rhetorical relation holds between the text segments and the photographs. Firstly, the GeM RST cannot bring the source of cohesion under analytical control, because the analysis does not extend below an orthographic sentence or a sentence fragment (see the list of base units in Table 3.1). For this dissertation, the benefits of the trade-off between the analytical granularity and the constraining notion of structure were discussed extensively in Chapter 3. Drawing a relation between the two segments would require loosening the analytical criteria, and thus work directly against the reasons for adopting the GeM model in the first place.

Secondly, if the text-flow were wrapped around the photographs, a rhetorical relation between the two might be more plausible. In this case, however, the photographs belong to different layout nodes in the hierarchy, as indicated by the bounding boxes in Figure 7.9. Moreover, the text and the photographs also occupy different areas in the area model.

What can then be said about the conceptual photographs without any additional concepts — such as cohesion — to describe their content and interaction with language? In fact, many useful observations can be made while working with the notion of structure. For example, we now know that the conceptual photographs are incorporated into the artefact structure with relatively weak rhetorical signalling. At the same time, Molina and Esteban (2006, p. 1045) have shown that the photographs have a considerable effect on forming an image of the destination.

Table 7.3: Photographs in the rhetorical structure

Identifier	Year	Photographs	RST relations
HFT	1984	28	6
HFT	1988	28	26
GNO	2008	11	1
HDB	1972	10	9
H45	2000	10	5
BNO	2006	10	1
HVG	2008	8	8
SSH	1986	7	7
HFT	1967	5	5
HDB	1969	5	5
TGH	1980	5	5
WEH	1995	5	4
WEH	1998	5	5
HFT	1972	4	4
HYW	1999	4	4
SMF	2001	4	1
WDH	2006	3	2
SHE	2002	2	1
WHE	2003	2	2

I would argue that the contrast between weak signalling and high impact of the conceptual photographs results from their established function in the genre of a tourist brochure. To put it simply: the photographs can convey an image of the destination, location or event without being tightly integrated in the rhetorical structure of the tourist brochure. In contrast, when the designers want to highlight specific visual content, this may be achieved by integrating the photographs into the multimodal structure as captioned photographs and illustrated descriptions.

An overview of the structural configurations involving photographs and their distribution across the data may be achieved by querying both layout and rhetorical layers of the GeM model. First, the realisation information component of the layout layer is used to identify the photographs, which are then matched with the corresponding RST segments in the rhetorical structure. The second step is to use the identified RST segments to find out how many RST spans the aforementioned segments participate in: I will clarify this issue below.

The results are shown in Table 7.3, which provides the identifier, the number of photographs and the number of multinuclear spans with a photograph as a

nucleus. Table 7.3 is listed in a descending order according to the number of photographs. Together, the number of photographs and associated RST spans can reveal the rhetorical configuration of the photographs. When a large number of photographs occurs in a few RST spans, the configuration in question is likely to be a cluster of conceptual photographs (*HFT* 1984). A close correspondence between the number of photographs and spans indicates that the photographs are in a direct relation with an RST segment. In many cases, these instances consist of the photographs and their captions (*HFT* 1988). In short, variation in the ratio between photographs and spans can be seen to reflect changes in the rhetoric-layout interface of the multimodal artefact (cf. Hiippala 2013).

Yet it is important to exercise caution, because Table 7.3 shows the results only for the brochure *series* and not for individual double-pages in the annotated corpus (see Section 5.1.3.2). For the fully annotated brochures, such as *HFT* 1984 and 1988, the ratio between photographs and RST relations can be used to correctly infer the integration of photographs into the rhetorical structure. However, in some series — such as *BNO* 2006 and *GNO* 2008 — a cover page with conceptual photographs may skew the results, although the content pages would use text-flow (see e.g. Figure 3.10).

For the most part, the method described above provides an effortless way to initially identify whether the photographs form image-text-complexes or provide a conceptual representation of the destination. The difference, however, only emerges when the number of photographs is sufficiently high. Therefore, the brochures with a low number of photographs, participating in a correspondingly low number of spans need to be subjected to a closer analysis. To conclude, inferring the correct rhetorical configuration requires a sufficient amount of data; otherwise a pattern cannot be detected.

To sum up the discussion of photographs, it can be said that they mainly participate in two types of structures in the tourist brochures. The first structure, an image-text-complex, provides a strong rhetorical relation between a photograph and the accompanying text. The structural simplicity of the image-text-complex allows its adaptation for a wide range of content types. In contrast to the image-text-complex, the second structure — a conceptual photograph — is rhetorically weak, but provides a powerful mechanism of representing the more abstract aspects of a destination. Lacking a strong, rhetorically signalled connection to other content in the artefact *structure*, conceptual photography is commonly used to convey an image of the destination and the values associated with it. Note, however, that this does not rule out the existence of cohesive ties between the photographs and text-flow, which are a common feature of semiosis in many genres.

This concludes the current discussion of the photographs and their functions and structure. I will discuss the structural configuration of photographs and text-

flow again in Section 8.1. I will now continue with other graphic elements, that is, the illustrations and maps.

## 7.3 Illustrations and maps

The illustrations, which constitute the third most common graphic element in the data (see Table 7.2), appear to have two distinct functions in the tourist brochures. The first function closely resembles that of the photographs: the illustrations provide a visual representation of the destination, location or event. The second function, in turn, is related to the subnuclear relation of IDENTIFICATION and its frequent appearance in connection with the illustrations. I will now discuss both of these functions below.

### 7.3.1 Illustrations as hand-drawn representations

Structurally, certain illustrations appear to serve a function similar to the photographs, that is, they are used to provide a representation of the destination to the reader. Again, these representations are mainly handled using the relation of RESTATEMENT. Like their photographic counterparts, the illustrations also participate in image-text-complexes, as exemplified by the illustration (s-2.01) and its caption (s-2.02) in Figure 7.10. According to the annotated corpus, the illustrations were most prominent during the 1970s.

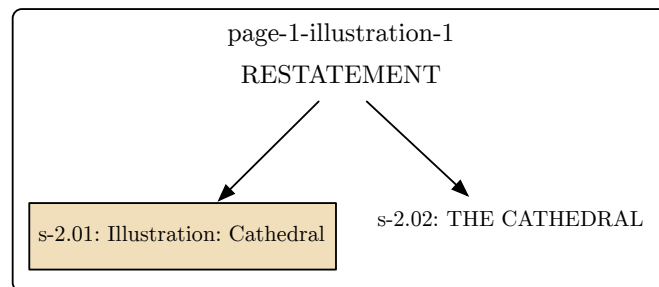
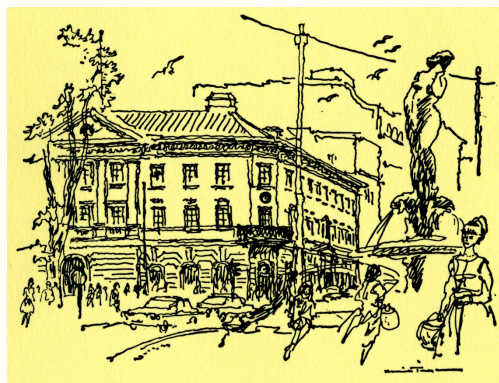


Figure 7.10: An illustration in an image-text-complex in *HDB* 1969

Although I expressed concerns about the concept of coding orientation in Section 2.4.2, the concept may be used point out the difference between illustrations and photographs (Kress and van Leeuwen 1996, pp. 170-171). Whereas the photographs provide a “naturalistic” representation with an apparently high truth value, the illustrations allow the designers to create a representation from scratch and according to their specific needs. It is therefore not surprising that the data includes both playful, cartoon-like illustrations (Figure 7.11a) and hand-drawn



(a) Illustration in *HFT* 1976



(b) Illustration in *SMH* 1982

Figure 7.11: Examples of illustrations

cityscapes (Figure 7.11b). In this connection, note that no vocabulary exists to describe or categorise these illustrations: they are currently an under-researched area in multimodal research (for emerging work in this area, see Guijarro and Pinar Sanz 2008; Astorga 2009; Guijarro 2011; Forceville 2011*b*; Boeriis and Holsanova 2012).

Nevertheless, the contemporary brochures rarely use illustrations to realise the different stages of the genre. Instead, they seem to complement the content realised using text-flow. This is exemplified in Figure 7.12, which shows an illustration associated with the header of a section. Together, they form a TITLE span for a list of suggested activities for the afternoon. The illustration, which (s-2.18) complements the header “1200-1600hrs” (s-2.19), shows the face of a clock with the aforementioned hours highlighted — a plausible case for a RESTATEMENT relation in the RST annotation.

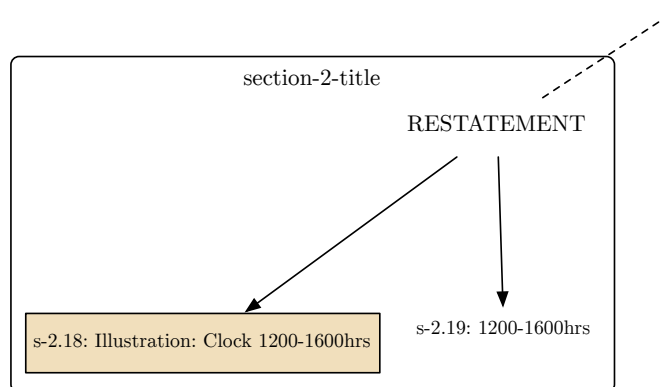


Figure 7.12: An illustration accompanying a header in *GNO* 2008

This leads to another issue: the two-dimensional elements. These elements, which account for roughly 30% of the graphic elements in the corpus (see Table 7.2), rarely contribute to the rhetorical structure (in only 8 instances out of 106). In most cases, the two-dimensional elements consist of geometric shapes, lines, arrows and so on, whose purpose is to connect and disconnect the content presented in the brochures. Consequently, I will not to dedicate a specific section for the two-dimensional elements, but simply state that their contribution to the rhetorical structure is not as significant as that of the illustrations. They play, however, an important part in organising the content in the semiotic mode of page-flow — an issue to which I will return in Section 8.2.1. Moreover, as I will show in the following section, certain illustrations are also involved in the rhetorical structure.

### 7.3.2 Illustrations and identification

Interestingly, a number of illustrations (n = 18) occur in rhetorical spans with the relation of IDENTIFICATION. The relation of IDENTIFICATION is a subnuclear relation, which is a relation intended to capture the rhetorical relations between segments that traditional RST would not acknowledge as analytical units (for subnuclear relations, see Section 3.2.3.2). A closer examination revealed two cases in which illustrations are used together with IDENTIFICATION; I will now describe both cases in detail.

The first case is associated with a specific graphic element — the Helsinki coat of arms — which often appears in connection with information about the publisher and whom to contact for more information (see Figure 7.13). As I have pointed out elsewhere (Hiippala 2012a, p. 111), a coat of arms is a highly codified symbol, yet its usefulness to the tourist is somewhat questionable. However, one explanation is that the coat of arms lends authority and credibility to the information in the brochure by signalling the official status of the publishers. As Pagani (2009) has showed, the state and other authorities often manifest themselves through

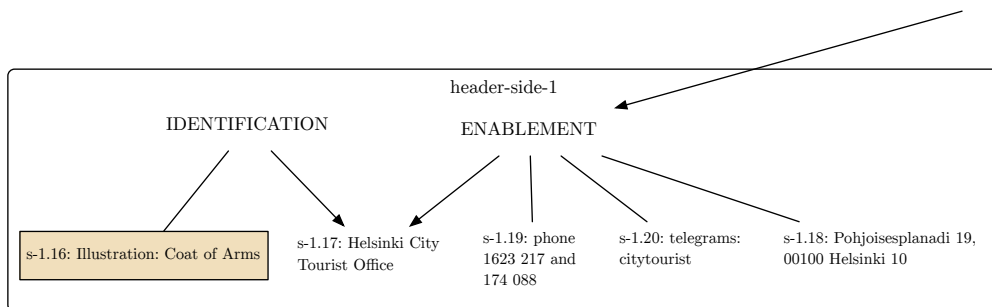


Figure 7.13: An illustration in an IDENTIFICATION relation in *HDB* 1972



multimodal symbols. This is a plausible explanation, as the coat of arms appear during the entire period represented in the data — it should be noted that these kinds of highly codified symbols are also resistant to the effect of the semogenic processes.

The second case is more interesting from the perspective of multimodal structure, because it shows how the illustrations may contribute to the reader’s navigation and use of the brochure. These illustrations, which I have previously described as a part of an implicit navigation structure (Hiippala 2012*b*, p. 1505), can be found in two brochures in the same series (*HFT* 1967, 1972). In terms of structure, the illustrations participating in an implicit navigation structure resemble the configuration described above in Figure 7.12, because these illustrations also occur close to a header.

In the case of Figure 7.14, an illustration with a green symbol is placed next to the header. The illustration has a navigational function since it appears elsewhere in the brochure in the same form and it has a distinct colour scheme. On the other side of the brochure, the same green symbol is used to indicate the described location on the map. This green symbol can be seen in Figure 6.5. As a rule of thumb, colour appears to have an important role in the implicit navigation structures. In this way, the brochures can use IDENTIFICATION to establish a subtle and implicit connection between text-flow, graphic elements and maps, that is, the output of different semiotic modes (see also Figure 3.5).

With the illustrations now covered, one graphic element remains to be discussed. This graphic element is the map, whose multimodal structure cannot be described within this dissertation to a degree that would do justice to its complexity (for an extensive description, see MacEachren 1995). Instead, the maps will be considered as constituents of the tourist brochures’ multimodal structure.

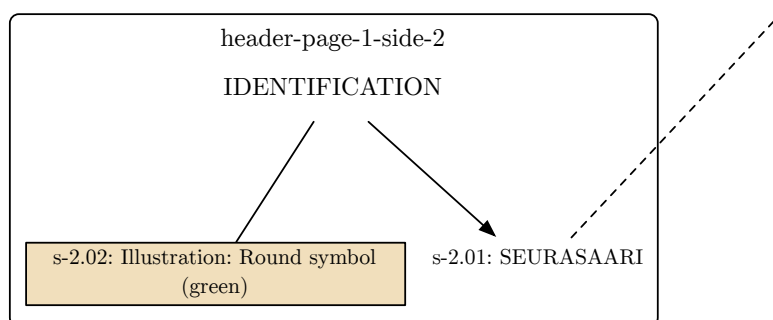


Figure 7.14: An illustration functioning as a navigation structure in *HFT* 1972

### 7.3.3 Maps

A common feature of the tourist brochures is a map, which helps the reader to navigate the destination. Various types of maps are used for this purpose: examples in the entire data set range from nautical to political maps. In the annotated corpus, maps can be found in 15 out of 30 brochures (50%). At the same time, maps may also be placed elsewhere than on the pages selected for the annotated corpus, so this number may not be representative of their actual use.

This brings us to the issue at hand. While a map is one of the most efficient ways of representing geographical and spatial information, it also takes up a considerable amount of layout space. A single map, linked to the content using a navigation structure, is often sufficient to serve the needs of a single brochure (for an example, see Figure 3.5). Moreover, maps rarely argue for their content: the most common rhetorical relation associated with them is the multinuclear JOINT. For this reason, I argue that to study how maps participate in the multimodal structure of the tourist brochures, the analytical attention should be directed towards the navigation layer.

Let us begin with an example from a brochure that guides the reader through a walking tour around two districts in Helsinki. In terms of the semiotic modes and rhetorical strategy, the brochure *Walking Tour in Helsinki* (1983) relies on the semiotic mode of text-flow in a sequential configuration (see Section 7.1.4). The brochure uses text-flow to lead the reader through the various locations, which are described in lesser or greater detail. No illustrations are used, except for a single instance to indicate the length of the walking tour. The most important graphic element is therefore the map, which shows the route and the sights situated along the walking tour.

To make use of the map, the text-flow constantly establishes links between the text and the accompanying map, as shown in Figure 7.15. These links may be captured using the navigation layer by describing their functions as pointers and entries. Overall, maps which function as entries in the navigation structure can be found in 7 out of 30 brochures in the annotated corpus.

To exemplify the use of a navigation structure, Figure 7.15 provides a visualisation of an explicit navigation structure. The visualisation was created by drawing a DOT graph on the basis of the data retrieved from the navigation layer. Figure 7.15 shows that a total of 32 numbered pointers in the text refer to the map: each box represents a single RST segment. In some cases, a single RST segment may contain up to three pointers; all of which point toward the same entry (see the three lines originating from **s-2.16** as opposed to a single line in **s-2.14**). In the tourist brochure, the numbers are highlighted typographically using a red background to make them stand out in the text-flow.

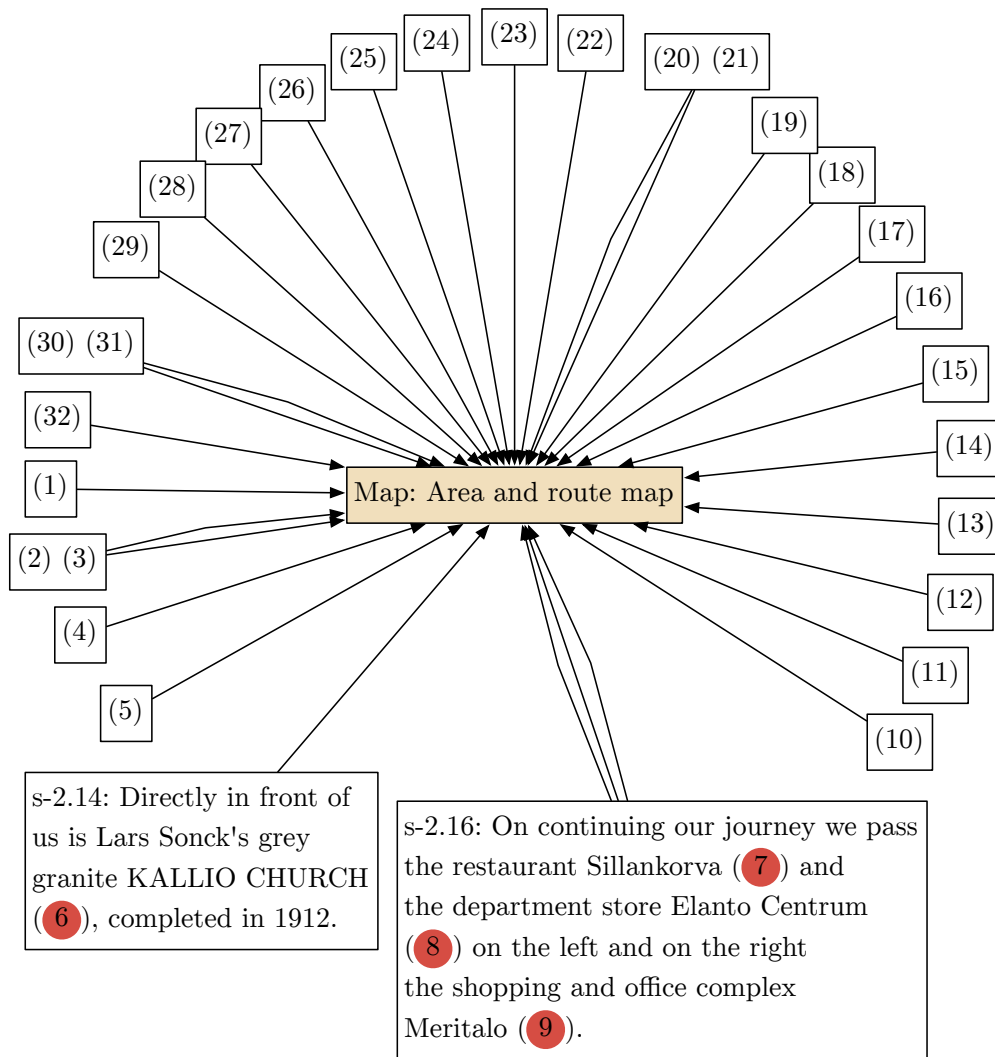


Figure 7.15: Navigation structure in *WTH* 1983

Figure 7.15 shows the power of navigation structures in multimodal artefacts, whose communicative function prompts the layout space to be used with maximum efficiency. A single map as an entry point for a navigation structure embedded into the rhetorical structure can bring together the content of an entire brochure. Obviously, the usefulness of navigation structures also depends on their easy use and understanding. A combination of a map — itself a highly complex visual representation of space — and a complex multimodal structure may result in a “cognitive overload” (cf. Mayer and Moreno 2003). Such an overload, whereby the meanings communicated using multiple semiotic modes exceed the cognitive

capabilities of the reader, may be avoided by clear and consistent signalling of the navigation structure.

The relationship between the maps and the multimodal structure of the artefacts that contain them could be explored further, but this falls outside the scope of this dissertation. To conclude, I have presented two possible avenues of research by identifying both implicit and explicit navigation structures. However, as the main goals of this dissertation lay elsewhere, I will now continue with a summary of the content described in this chapter.

## 7.4 Summary

This section provides a brief summary of the content, its structural features and role in the genre of a tourist brochure. In Section 7.1, I began the analysis with text-flow, which is a semiotic mode that can realise all stages of the genre. Text-flow provides the rhetorical strategies for describing the destination, location or event either in small or in great detail, depending on the communicative goals of the brochure. It also enables the reader to perform the role of a tourist by suggesting activities to do and routes to follow. For each aforementioned task, the text-flow adopts a different configuration, which can be captured using the rhetorical and layout layers.

A visual description of the destination, location or event, in turn, is provided using photographs and illustrations. In Section 7.2, I observed that the photographs often appear in two configurations: (1) image-text-complexes and (2) conceptual photographs. In image-text-complexes the hierarchical layout structure works together with the rhetorical structure to establish a strong link between a photograph and its caption. The alternative configuration — which I termed conceptual photograph — opts for weak rhetorical signalling and rarely establishes a direct connection with the linguistic content.

Illustrations are used to complement the linguistic content, as described in Section 7.3. As opposed to the naturalistic representations in photography, they allow an alternative coding orientation that endows the designers with more artistic freedom: the illustrations may be completely adapted to the needs of the brochure. In certain cases, the illustrations perform other functions as well: the illustrations provide a visual representation or contribute to the navigation structure. In the navigation structure, the illustrations connect different content by functioning as implicit pointers and entries. These implicit pointers and entries rely on visual similarity — typically based on colour and shape — for their recognition.

Finally, for the actual navigation in the destination, maps are indispensable for an effective representation of geographical and spatial information. However, as Section 7.3.3 pointed out, the maps take up a considerable amount of layout

space, which is often limited in the brochures. Therefore, a navigation structure — in the GeM sense — is used to connect the output of different semiotic modes. This is done using explicit pointers and entries, such as numbers.

To conclude, in this chapter I attempted to provide a detailed picture of the content of the tourist brochures, and in particular, the structure of this content. As the content and its structural characteristics were laid out above, I will now make explicit how the analyses performed up to this point relate to the overall goals of this dissertation.

So far, this dissertation has provided an extensive description of the brochure as a medium in Chapter 6. This chapter covered the various properties of the medium: the semiotic modes deployed, the relationship between advertising and a medium, and the impact of desktop publishing. In relation to multimodal analysis, the covered topics were very broad, yet an understanding of these issues was deemed necessary for the analysis to proceed.

The next step taken here in Chapter 7 was concerned with investigating the multimodal structure of the tourist brochures in great detail. I would argue that the analysed artefact has now been covered from two vantage points. Firstly, from a very broad perspective, and secondly, from a very detailed perspective. What remains to be covered is the middle ground between these two perspectives. This middle ground is the page. For this reason, in the following Chapter 8, I take a step back from analysing the content and instead focus the analytical attention on the multimodal structure of a page.

# Chapter 8

## The page and its interpretation

The final chapter of this dissertation develops a particular theme, that is, how the semiotic modes are combined on the pages of a tourist brochure. As Section 2.5 established, the page has attracted a fair amount of attention in multimodal research. For instance, Baldry and Thibault have suggested that “in modern society the *page* is an important textual unit”, whose recognition is “clearly reflected in the growing list of expressions that identify the page in terms of different social functions” (2005, pp. 57-58). If this observation holds true, an increased effort should be directed towards describing how different pages are construed multimodally.

The field of information design has also acknowledged the importance of the page and underlined its many implications to communication. In a recent paper, Waller argues that “page layout is a little-discussed aspect of text, but it connects closely to a range of fundamental issues concerning the nature of text, documents, writing and reading” (2012, p. 237). While Waller touches upon an important issue, that is, how the page both supports and influences the reading process, I will consider the page from the perspective of the semiotic modes and their discourse semantic interpretation.

Given the many forms of a page, a model of artefact structure has to be able to explain how the pages are construed in terms of structure and function, that is, what they do and how. It is also reasonable to expect that the tourist brochures include different page types — such as cover pages and content pages (see Section 2.5) — and as I will show, the developed framework can already bring the different page types under a considerable degree of analytical control by deploying the GeM model. With the annotated corpus at hand, the dissertation can provide a more informed view of the semiotic modes and their configuration. I shall now begin with a discussion of the challenges faced in the process of identifying the deployed semiotic mode.

## 8.1 The step from text-flow to page-flow

In Chapter 7, I used the GeM-annotated corpus to examine the multimodal structure of the tourist brochure in great detail. For instance, the analysis identified several configurations of text-flow, which were used to realise the stages of the genre of a tourist brochure. In addition, I also considered the integration of graphic elements into the multimodal structure of a tourist brochure.

However, what has not been covered yet is the semiotic mode of *page-flow*, which combines both text-flow and graphic elements in its expression (see Section 3.5). At this point, the defining characteristic of page-flow needs to be brought back into mind: the semiotic mode of page-flow combines the ‘output’ of the participating semiotic modes on the two-dimensional space of a page and uses the discourse semantics to signal the relations between the different modes (Bateman 2009b, 2011). In order to shed light on how different semiotic modes operate together as a part of page-flow, I will now move away from the detailed analyses of multimodal structure towards a more generic view by focusing on the page.

To grasp the role of page-flow in the tourist brochures, it is necessary to take a step back from the content and look at the entire page. I propose that a better understanding of page-flow may be achieved by bringing the perspectives of document theory described in Section 4.4 to bear on the data. In particular, the observation of Bateman and Schmidt that a multimodal artefact consist of “articulated parts designed to be put together in various ways by its users” (2012, p. 48) is enlightening. What we need to know then is what triggers the discourse semantic interpretation associated with page-flow, instead of that associated with text-flow. In terms of Waller (2012), we need to know which features of a multimodal artefact encourage a *non-linear* reading strategy (see Section 4.4.2).

In the analysis, I will place special emphasis on cross-layer analyses between the layout and rhetorical layers (Hiippala 2013). I predict that these layers contribute to the reader’s recognition of page-flow, because they shape the “articulated parts” of an artefact, which the users are expected to put back together during the process of interpretation (Bateman and Schmidt 2012, p. 48). To proceed, the next step is to establish the difference between text-flow and page-flow. And as I will show below, this difference has to be grounded in the notion of multimodal structure.

When faced with an overtly visual multimodal artefact, the analyst may be tempted to consider that the active semiotic mode is page-flow. To exemplify, according to the calculations performed for Figure 6.4, roughly 62% of the total layout space in *Welcome To Helsinki* (1998) is occupied by five photographs, which extend across the double-page (see Figure 8.1). The placement of these photographs is indicated using the identifiers **s-3.19** to **s-3.23** in Figure 8.2. As the area model in Figure 8.2 shows, the photographs are not placed *within* text-flow, but adjacent to it: on the left, on the right, and above. We can then present the

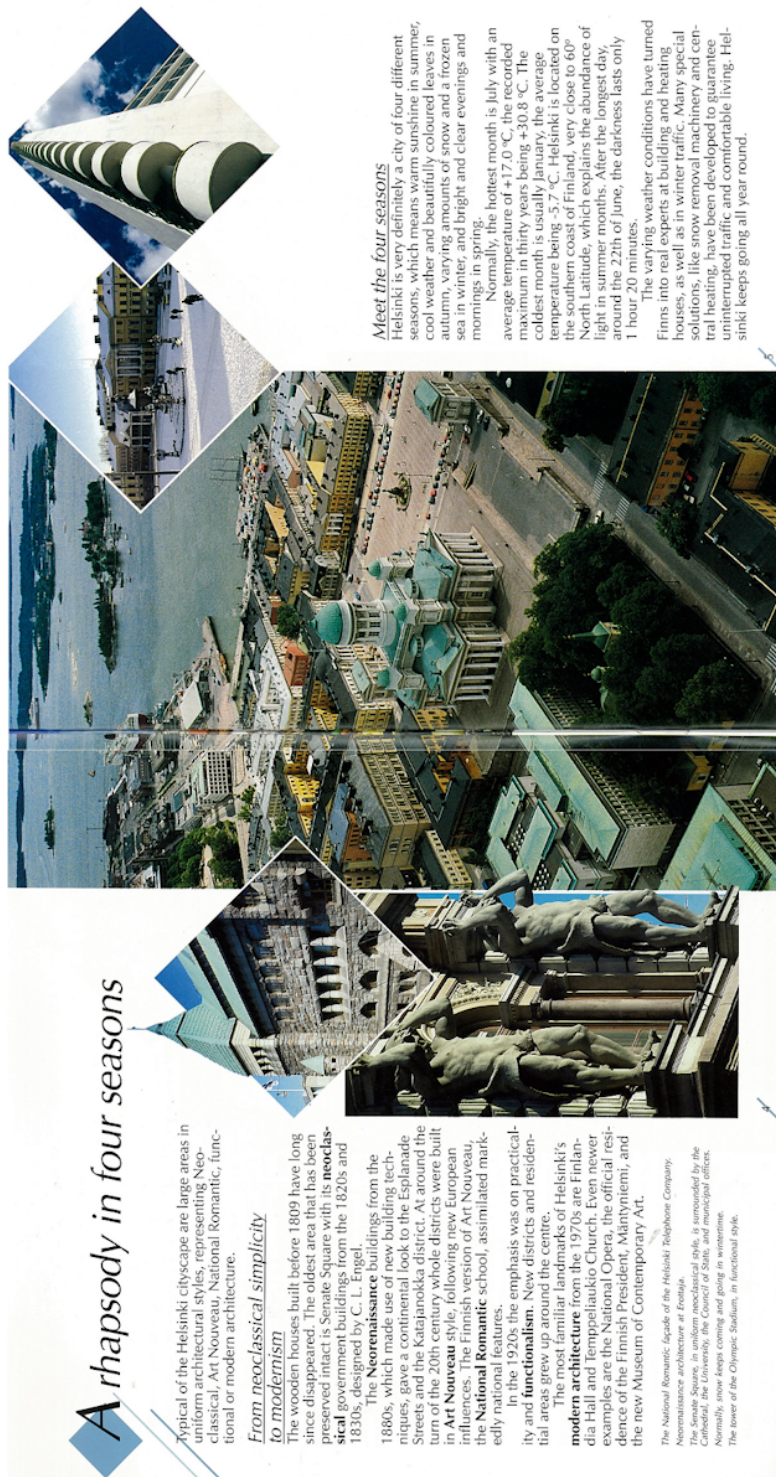


Figure 8.1: *Welcome To Helsinki* (1998)



following question: is this double-page an example of page-flow — or text-flow accompanied by an image-text-complex with a special configuration?

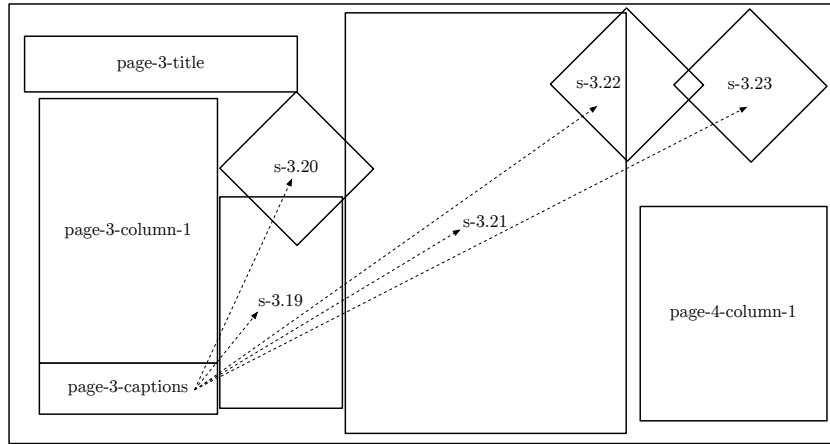


Figure 8.2: An area model of *WEH* 1998

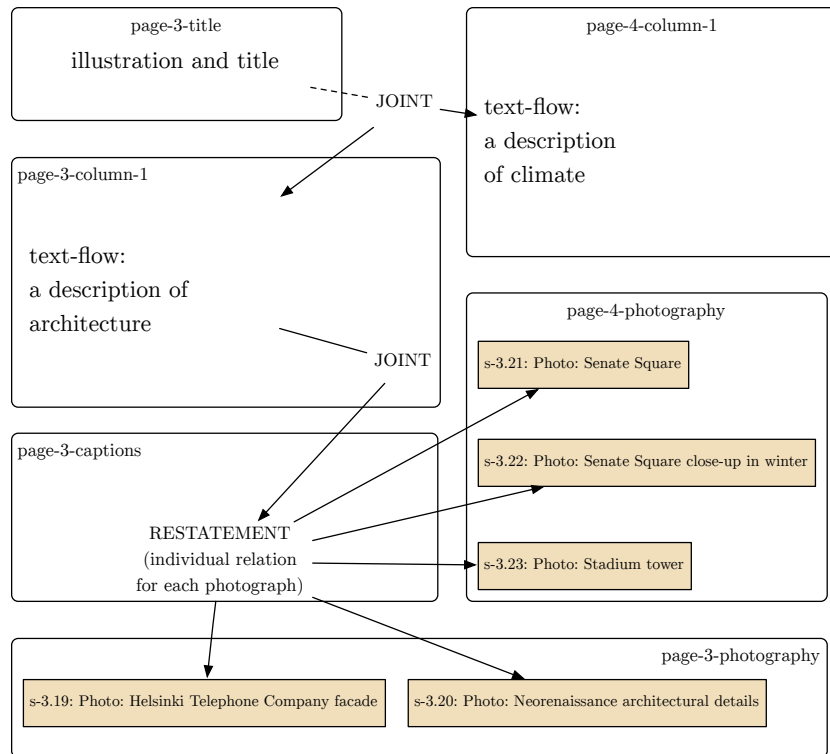


Figure 8.3: The layout and rhetorical structures in *WEH* 1998

Instead of making a decision based on the visual appearance of the double-page, a look at the rhetorical and layout structures may inform the identification of the active semiotic mode and strengthen its basis considerably. To begin with, the bounding boxes in Figure 8.3 show the layout chunks that constitute the layout structure. Additionally, Figure 8.3 shows the rhetorical relations that hold between the chunks. For increased legibility, the rhetorical structures of the text-flow columns have been collapsed. It should also be noted that a detailed analysis of the rhetorical structure in the layout chunk **page-4-column-1** was already presented in Figure 7.2.

As Figure 8.3 shows, the overall rhetorical structure of the brochure is typical of text-flow, apart from one exception. Curiously, all captions are grouped under the same layout chunk and placed in the same layout area (**page-3-captions**). For this reason, the photographs and their captions do not form distinct image-text-complexes (see Section 7.2.1).

To correctly infer the configuration of the image-text-relations in Figure 8.3, the reader must proceed through the captions and associate them with the correct photographs, which requires two distinct interpretations. For **page-3-photography**, the correct order of captions and photographs proceeds from top to bottom. In the case of **page-4-photography**, the correct order follows a path from left to right. Otherwise, the double-page bears the hallmarks of text-flow with an accompanying image-text-complex, but as the analysis shows, the interpretation of the page requires more than just a linear interpretation of text-flow.

In contrast, inferring the correct relation between the captions and the photographs involves making sense of their organisation on the page. I would argue that this speaks in favour of the decision that the active semiotic mode is page-flow. What is even more important, the interface between the rhetorical and layout structures seems to hold the key to understanding the principles behind page-flow (cf. Hiippala 2013). Thus, I will take this interface as the point of departure for the following investigation of page-flow.

## 8.2 Investigating page-flow

The first step in taking page-flow apart in analysis involves a consideration of the semiotic modes at play. As I already pointed out in Section 8.1, the difference between ‘visually-enriched’ text-flow and page-flow is not always clear, because the degree of visuality alone is not sufficient to justify that the active semiotic mode is page-flow. The only way to bring out the difference between the semiotic modes is to dig deeper into the structure of a multimodal artefact. At this point, it is also useful to call to mind the definition of page-flow in Bateman (2011, p. 26):

[Page-flow] relies upon the complete two-dimensional space of the ‘canvas’ provided by the physical substrate and uses proximity, grouping of elements, framing and other visual perceptual resources in order to construct patterns of connections, similarity and difference.

The quote above provides an appropriate starting point for deconstructing page-flow, but it should be kept in mind that the process of visual perception (Kappas and Olk 2008) and the presence of visual perceptual resources on a page are fundamentally different but interrelated issues (Leckner 2012). Here, I focus mainly on the role of visual perceptual resources in the organisation of the content.

Armed with the detailed descriptions of the content and its multimodal structure in Chapter 7, the dissertation now stands in a good position to take on their combinations on a page, as we are now more aware of how the structures work towards certain communicative goals. The next question is: does the organisation of the content have implications for the interpretation of the page?

As Section 3.3 established, the discourse semantic component guides the interpretation of the semiotic resources that participate in the semiotic mode. The discourse semantics also play an important role in the interpretation of a page: Bateman (2011, pp. 33-34) has shown that making sense of a single page may require applying multiple discourse semantic interpretations. Obviously, the processes of interpretation involving an entire page are also of high interest to a model of a multimodal artefact. To move forward, I will now direct the analytical attention towards the interface between the layout structure and the rhetorical structure.


### 8.2.1 Interrogating the layout structure

I shall begin with an analysis of a leaflet describing the 18<sup>th</sup> century fortress of Suomenlinna, a UNESCO World Heritage Site and a major tourist attraction located on the islands in front of Helsinki. The leaflet, published in 1986, is entitled *Suomenlinna Seafortress in Helsinki*. In the annotated corpus, the leaflet was assigned the identifier *SSH 1986*: I shall refer to the leaflet using this identifier (see Table 5.2).

In terms of the fold geometry, the leaflet is two-sided with three folding points. The side of the leaflet with the most content is illustrated in Figure 8.4: I shall refer to it as the content side. The calculations performed for Figure 6.4 show that 13% of the layout area is taken up by visual elements. A quick look at Figure 8.4 also reveals that the leaflet combines several types of content discussed in Chapter 7: paragraphed text-flow, image-text-complexes and a colour-coded table.

In terms of the “visual perceptual resources”, the leaflet matches several characteristics of page-flow outlined in Section 8.2 (cf. Bateman 2011, p. 26). The

### Piper Café



The café stands on Suomenlinna's highest point and commands a splendid view of the sea and the surrounding islands. There are also tables outside. Open 10 May—31 August, 10 am–5 pm.

### Helga Café Ship

Converted sand yacht in Tyköslahdi Bay. Food also served. Specialty pizzas. Open from 1 May Mon.—Sat 10 am–8 pm, Sun 10 am–6 pm.

### Katharina Café

Located in the Nordic Arts Centre. Opening hours same as for the Jerry Barracks Gallery.

### Public beach

The beach, on the northern cove between Suissaari and Kustaanmiekka islands, is open and guarded 15 June—16 August daily 10 am–6 pm. Shower and changing facilities.

### Summer theatre

Plays are performed in Finnish at the Good Conscience Bastion during the summer months. Information, tel. 718 022.

### Information kiosk

The kiosk alongside Tyköslahdi Bay sells brochures, maps, souvenirs etc. Guided tours also start here. Tel. 668 341.

### Boat connections

1. FERRY, MS Suomenlinna-Sveaborg, Market Square—Suomenlinna, Jerry-Bastion, 1 May—30 November Sat—Sun.  
 2. FERRY, MS Suomenlinna-Sveaborg, Market Square—Suomenlinna, Jerry-Bastion, 1 May—30 November Sat—Sun.  
 3. WATERBUSSES, MS Amiralai, MS Walhalla and MS Vapouri. Market Square—Tyköslahdi Bay—King's Gate—Market Square. 17 May—31 August, 10 am–5 pm.  
 4. WATERBUSSES, MS Amiralai, MS Walhalla and MS Vapouri. 1 May—16 June and 1—14 September. Open except Sundays and Sundays \*\* - last boat on Sundays and wharf restaurant is closed.

		SUOMENLINNA—HELSINKI			
		1	2	3	4
HELSINKI—SUOMENLINNA		7:40	8:00*	8:00*	7:30*
		8:30	9:00	9:00	8:30*
		9:30	10:00	10:00	9:30
		10:30	11:00	11:00	10:30
		11:30	12:00	12:00	11:30
		12:30	13:00	13:00	12:30
		13:30	14:00	14:00	13:30
		14:30	15:00	15:00	14:30
		15:30	16:00	16:00	15:30
		16:35	17:00	17:00	16:30
		17:30	18:00	18:00	17:30
		18:30	19:00	19:00	18:00
		19:30	20:00	20:00	19:30**
		20:30	21:00	21:00	20:30
		21:30	22:00	22:00	21:30
		22:30	23:00	23:00	22:30
		23:30	24:00	24:00	23:30
		0:30	1:00	1:00	0:30
		1:30	1:30	1:45	1:45

### Tourist information in Helsinki

Information revised January 1986  
 Helsinki City Tourist Office  
 Pöyhönenkatä 19  
 Tel. 169 3737  
 The publisher is not responsible for possible changes.  
 Publisher: Helsinki City Tourist Office in co-operation with the Ehrensvärd Society.

### Guide services

The best way to find out about Suomenlinna is to take one of the guided tours offered daily. Tours start from the information kiosk at Tyköslahdi Bay and last about 1½ hours. Your guide will point out places of interest in the fort and will also take you to the islands of the fort.

### Ehrensvärd Museum

Historical museum on Suissaari Island. It includes a model ship collection and officer's quarters dating from the 18th century, as well as exhibitions detailing Suomenlinna's military past.

### Armfelt Museum

Reconstruction of a 19th century upper-class home in a former barracks. The museum includes five well-furnished rooms and a collection of china. Model ship exhibition in the summer of 1986. Open daily 12 am–5, 30 pm. In September Sat–Sun only.

### Coastal Artillery Museum

Arms museum in a vaulted second world war bunker. Open daily 11 am–5 pm, 31 August daily 11 am–5 pm.

### Submarine Vesikko

Situated on the northern eastern edge of Suissaari Island. The Vesikko was built in Turku in 1931–33 and took part in World War Two. Visitors are allowed on board. Open 10 May—31 August daily 11 am–5 pm.

### Nordic Arts Centre

Exhibitions and book sales at the Jerry Barracks Gallery and the Garrison Barracks on Suissaari Island.

### Suomenlinna Church

Orthodox garrison church dating from 1874 which has been redecorated in the Lutheran style. Open Sundays 12—4 pm.

### Walhalla Restaurant

Located on Kustaanmiekka Island, the Walhalla includes a first-class restaurant, cafeteria and pub. A brick casemate dating from the 1790s serves as a kitchen. The terrace is open from 10 am–11 pm on Saturdays and Sundays. Tel. 668 552.

Figure 8.4: The content side in *SSH* 1986

content is grouped into a total of four columns on the content side. Moreover, each column realises stages of the genre of a tourist brochure, which mainly take the form of location and event descriptions (see Section 4.3.1). As I will show below, these descriptions follow a consistent structural pattern, which closely resembles what Waller et al. (2012) have described using the notion of a pattern language (see Section 4.4.3). In short, a pattern provides a solution to a communicative need by providing the required means of expression. These patterns are the “building-blocks” of a genre, which Stöckl (2004) conceptualised in his work on modal interrelations (see Section 2.3.1.3).

The structures identified in Chapter 7 exemplify the patterns necessary for realising the genre of a tourist brochure. The key to identifying these patterns was the co-deployment of the GeM model and the notion of the semiotic modes. As I will show below, the same analytic tools can take us further in the analysis of entire pages. In the case of Figure 8.4, I would argue that the active semiotic mode is page-flow, but to understand how page-flow operates, the multimodal structure of the leaflet needs to be brought under closer analytical attention. To move forward, I propose that special attention should be paid to the layout structure. Note, however, that owing to limited space, the layout structure in Figure 8.5 only covers the fourth column on the right in Figure 8.4.

First of all, two image-text-complexes may be identified in the layout structure. Both are marked using a grey background in Figure 8.5. The white boxes, in turn, indicate layout leafs with text-flow. For a closer look at the rhetorical structure of the image-text-complex with the identifier **image-text-complex-1**, see Section 7.2.1. In Section 7.2.1, I argued that the image-text-complex relies on the spatial proximity of the semiotic resources to signal that they are to be interpreted together. But to consider how the image-text-complexes work as a part of a page, it is necessary to take a step back in the analysis.

There are a total of six layout chunks in Figure 8.5. In addition to the two image-text-complexes, there are four layout chunks that consist entirely of text-flow. To realise the genre of a tourist brochure, each text-flow chunk uses the same structural pattern, which may be captured using the GeM model. In each case, a typographically emphasised header is followed by two text columns, which are separated by a demarcating vertical line. Moreover, several horizontal lines — situated directly under the parent node (**column-4**) — are used to separate the text-flow chunks and image-text-complexes in Figure 8.5. While the layout layer can capture the visual perceptual resources present on the page, it is the rhetorical layer which allows us to describe how the content is presented and argued for.

In this context, the key to understanding how page-flow operates is to acknowledge that the layout structure signals that *rhetorically*, the content is to be interpreted separately, chunk by chunk. If multimodal artefacts indeed consist of

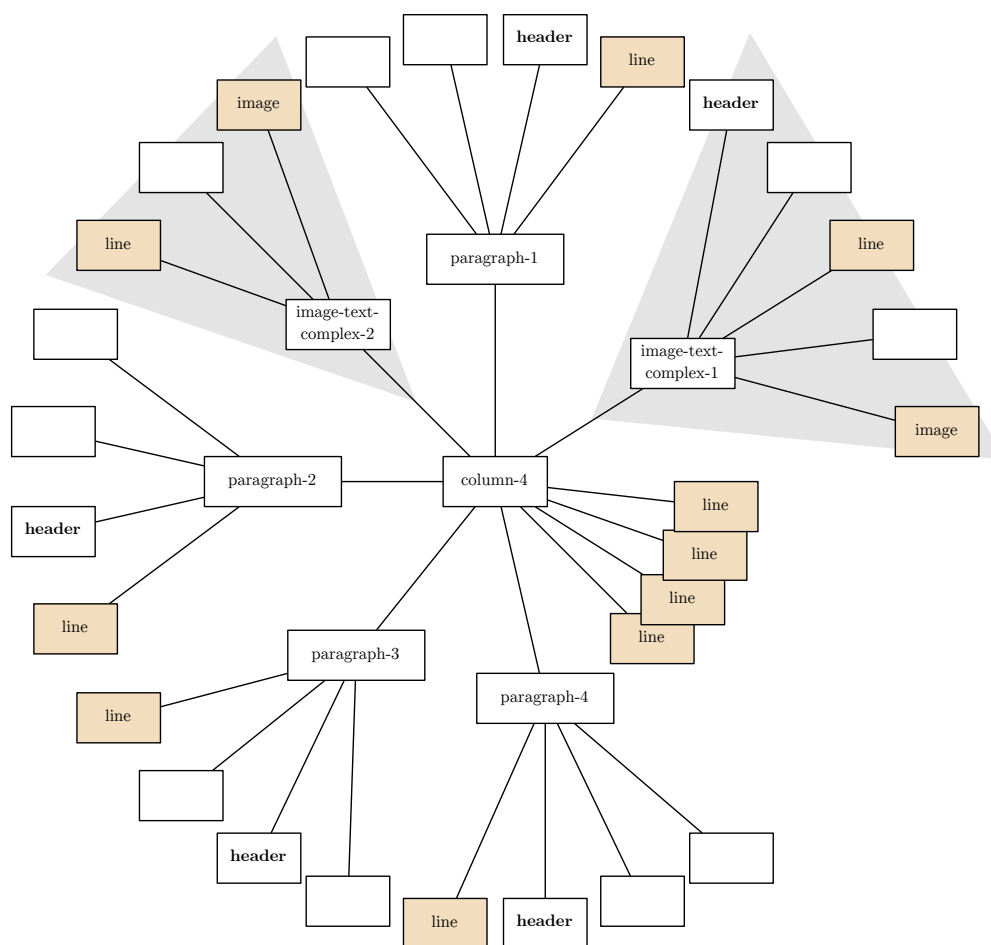


Figure 8.5: The layout structure of the fourth column in Figure 8.4

articulated parts that the users put back together during the interpretation process (Bateman and Schmidt 2012, p. 48), then I would argue that by organising the content under separate layout chunks, the layout structure in Figure 8.5 strongly encourages the adoption of a selective reading strategy (see Section 4.4.2) (cf. also Waller 2012, p. 239). This kind of reading strategy supports the kind of discourse semantic interpretation associated with page-flow.

A completely different interpretation might take place upon encountering another layout structure, which nevertheless realises the same genre of a tourist brochure. Such a structure may be found on the other side of the same brochure, which contains the cover of the leaflet and a description of Suomenlinna as a destination (see Figure 8.6). From now on, I shall refer to this side as the cover side. As opposed to the content side, the cover side is mainly visual: 87% of the layout space is occupied by graphic elements (cf. Figure 6.4). The layout structure of



Figure 8.6: The cover side in *SSH* 1986

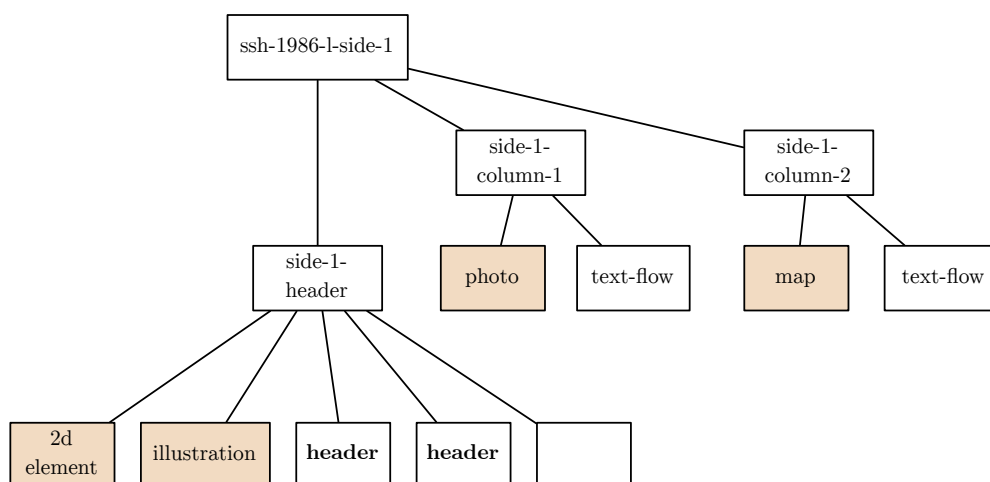


Figure 8.7: The layout structure of side one in *SSH* 1986

the entire cover side is represented in Figure 8.7. In contrast to the example of page-flow in Figure 8.5, I propose that Figure 8.7 exemplifies a layout structure that is consistent with the semiotic mode of text-flow.

I base my argument on the following observations. Firstly, unlike the highly fragmented structure of the page-flow, the layout structure shown in Figure 8.7 is relatively simple with only three layout chunks for the entire page. The layout chunks that carry the actual content (**side-1-column-1** and **side-1-column-2**) both contain two child nodes. In both chunks, a layout leaf with text-flow is accompanied either by a photograph or a map. Moreover, no demarcating lines are present in these layout chunks. Secondly, the layout structure contains only a single layout chunk with two headers (**side-1-header**), both placed on the brochure cover. The content side, in contrast, contains a total of six headers, which help the reading process by organising the content: I argue that this organisation is consistent with the semiotic mode of page-flow.

To further emphasise the difference between the two layout structures, Table 8.1 quantifies the layout structures in the *Suomenlinna Seafortress in Helsinki* leaflet. Whereas the layout structure of the cover side contains a total of three chunks and nine layout leaves (see Figure 8.7), the corresponding structure on the content side consists of 36 layout chunks and 154 layout leaves. The high number of layout leaves may be explained by the presence of the colour-coded table and its many cells (cf. Figure 7.1).

In addition, the layout structures on the two sides are of different depth. In the case of the content side, a move through the layout structure from the topmost parent node to the last layout leaf on the bottom requires a total of six steps. On the cover side, the same process requires only three steps. On the basis of Table



Table 8.1: Layout chunks and layout leafs in *SSH* 1986

	Chunks	Leafs	Depth
Cover side	3	9	3
Content side	36	154	6

8.1, the layout structures appear to differ in terms of the number of layout chunks and leafs, and the depth of the layout structure. A plausible explanation for this difference is provided by the deployment of different semiotic modes.

I argued above that two different semiotic modes are active in the leaflet: text-flow on the cover side and page-flow on the content side. At this point, the question that arises is whether the active semiotic mode on either side may be determined by the layout structure alone? While the layout structure may help us to identify the semiotic mode, it needs to be kept in mind that the layout structure is not arbitrary. Instead, this structure is functionally motivated, and for this reason the rhetorical structure needs to be brought in to complement the analysis. In this way, we can describe how the different “articulated parts” on a page come together via the layout structure (Bateman and Schmidt 2012, p. 48). This provides the point of departure for the following section, which examines the layout structures in the annotated corpus in greater detail.

## 8.2.2 Layout structures in the annotated corpus

Table 8.2 lists the 15 pages with the highest number of layout chunks in the annotated corpus, sorted in a descending order. Additionally, Table 8.2 provides the number of layout leafs and the percentage of layout space occupied by visual elements for each page. The leaflets are marked using a grey background; the white background indicates a brochure (for brochures and leaflets, see Section 6.1.2).

The first five pages in Table 8.2 contain a high number of layout chunks, yet they do not include any photographs, illustrations or maps. This results from the fact that these pages are either table- or list-driven. As I pointed out in Section 7.1.1, the tables and lists draw on the hierarchical layout structure for their meaning potential. In principle, their structure may be characterised as *fragmented*: the bits and pieces of meaning are brought together in the layout structure and made accessible by their systematic visual organisation into the rows and columns of a table or into a list of items (cf. Lemke 1998; Baldry and Thibault 2005).

Having already discussed the structure of tables in Section 7.1.1, I will now proceed to consider another page listed in Table 8.2, which is positioned right after the leaflet discussed in Section 8.2.1 (**ssh-1986-1-side-2**). This page from



Old Town Rapids



Power Station Museum



Compass Square (Kompassiaukio)

## Other interesting sights and attractions

### Old Town

Helsinki was originally established at the mouth of the Vantaanjoki River in 1550 by King Gustav Vasa of Sweden. Today this area is one of Finland's most modern centres for the arts, sciences and high technology. Contemporary architecture can be admired in the new Arabianranta residential areas along the shores of the Old Town Bay. The district also houses the distinguished University of Art and Design Helsinki, which attracts students from around the world.

To get to the Old Town and Arabianranta, take Tram 6 or Bus 68, 71, 71V or 738 from the Railway Station Square in the city centre.

### Hakaniemi Market



### Arabia Museum and Factory, 38

Arabia has been a pioneer in modern Scandinavian design, combining timeless beauty with everyday practicality. The porcelain factory has manufactured household and art ceramics in the same location for over 130 years.

Hämeentie 135,  
Tel. +358 (0)204 38 5357.

**Arabia Museum** open: Tue-Fri 12 noon-6pm, Sat-Sun 10am-4pm. Entrance fees: adults €6, children (7-16) €3 and special groups €1.50/€0.90 per person.  
[www.arabiamuseum.fi](http://www.arabiamuseum.fi)

### Factory tours: reservations

Tel. +358 (0)204 38 5326. Guided tours are available for 10 persons or more. **Arabia Outlet**, **Finlayson** and **Perintä** factory outlets are open: Mon-Fri 10am-8pm, Sat-Sun 10am-4pm.

The **Technical Museum** alongside the Old Town Bay presents the achievements of old and new technology and hosts related exhibitions.

Vikintie 1, tel. +358 (0)9 7284 4428. Open: Wed-Sun 12 noon-4pm, see [www.technicalmuseum.fi](http://www.technicalmuseum.fi) for fees: adults €6, children (7-16) €1, pensioners and unemployed €5, students €2.



Nearby you can also find the **Power Station Museum**, featuring original hydroelectric pumps from the 1800s.

Hämeentie 163, tel. +358 (0)9 3108 7064. Open summer time: Wed-Sun 1am-5pm, see [www.helsinginkiaupunginmuseo.fi](http://www.helsinginkiaupunginmuseo.fi)

### Hakaniemi District 39

#### Hakaniemi Market

Groceries, flowers and baskets have been sold at Hakaniemi Market for over a hundred years. While shopping you can enjoy coffee at one of the outdoor cafes.

Opening hours: Mon-Sat 6.30am-3pm, plus on the first Sunday of each month from 10am-4pm.

#### Hakaniemi Market Hall

Hakaniemi Market Hall is located beside the outdoor market. Upstairs you can find a variety of clothing, souvenirs and designer products while downstairs you can buy groceries and delicacies.

Opening hours: Mon-Fri 8am-6pm, Sat 8am-4pm.



Kaivopuisto Park

**Kaivopuisto Park** is Helsinki's oldest and most famous park and is overlooked by 900-year-old castles. The park also has many embassies, including those of the USA and Russia. The proximity of the sea, rock cliffs and green lawns offer many opportunities to relax and enjoy the outdoors. At the highest point of the park is the **Ursa Observatory**. There are several cafes and restaurants along the shore and on the nearby islands.

### Sinebrychoff Art Museum 40

The Sinebrychoff Art Museum is specialised in old European art and includes Finland's most important collection of old masters. The art collections include especially portraits and artefacts from the 17th and 18th centuries. The upper floor is a wonderful reproduction of the family home of the famous brewer.

The museum presents its collections through exhibitions, educational programmes. The Sinebrychoff Art Museum is part of the Finnish National Gallery.  
Bulevardi 40, tel. +358 (0)9 1733 6460. Opening hours: Tue and Fri 10am-6pm, Sat 12 noon-6pm, Sun 12 noon-5pm. Entrance fees: adults €4.50/3 or €7.50/6, visitors under 18 free. Free admission on Thursdays 5pm-8pm. [www.sinebrychoffinmuseo.fi](http://www.sinebrychoffinmuseo.fi)

### Hakaniemi District

Hakaniemi Market is a popular place to shop, especially as it also includes a year-round open-air flea market



Merisatamantien shoreline

(weather permitting). The market offers an enormous variety of things to buy, as well as a friendly and relaxed atmosphere.

Tel. +358 (0)9 310 71396. Hietalahti Market is open Mon-Fri 6.30am-6pm, Sat 6.30am-4pm. The flea market is open on Wednesdays and Saturdays. There is also a **Sunday flea market** open May to September from 10am-4pm.

### Hietalahti Antique and Art Hall

Hall is located alongside and sells a wide range of furniture, jewellery and other antique items.

Opening hours: Mon-Fri 10am-5pm, Sat 10am-3pm.  
[www.hietalahdenkauppaalli.fi](http://www.hietalahdenkauppaalli.fi)

### Kauppalahteen (Cable Factory) 41

The massive **Cable Factory** was used up until the 1950s to produce telegraph and electric cables. Today the building is a centre for culture and the arts. Over 800 artists, artisans and other cultural people work within these walls. The Cable Factory houses art schools, dance theatres, two radio stations, three museums and eight galleries. There are also many

special groups €2.50. [www.teatterimuseo.fi](http://www.teatterimuseo.fi)

### Hotel and Restaurant Museum

15 free, special groups €7/person. [www.hoteljavoinnolmuseo.fi](http://www.hoteljavoinnolmuseo.fi)



Ruoholahti Canal

varieties of sports clubs. The Cable Factory is also home to the popular cafe and restaurant Hietalahti. See also page 46.

### Tallberginkatu 1 C

Tel. +358 (0)9 4763 8330. Opening hours: (museum): Tue-Sun 11am-6pm. [www.kappeliellat.fi](http://www.kappeliellat.fi)

### Cable Factory museums

**Finish Museum of Photography**  
Tel. +358 (0)9 6866 3622. Entrance fees: adults €6; visitors under 18 free; special groups €4, groups (min. 25) person. [www.fmp.fi](http://www.fmp.fi)

### Theatre Museum

Tel. +358 (0)207 961 670. Entrance fees: adults €5.50, free, children under 13 and special groups €2.50. [www.teatterimuseo.fi](http://www.teatterimuseo.fi)



Ruoholahti and Kauppalahteen

Figure 8.8: A double-page in *Helsinki Visitors Guide* (2008)

Table 8.2: Pages with the highest number of layout chunks

Identifier	Chunks	Leafs	Graphics
ceh-1967-l-side-2	153	271	0 %
ceh-1967-l-side-1	110	185	0 %
bno-2006-l-side-2	48	71	0 %
meh-1967-l-side-2	38	172	0 %
smh-1982-l-pages-4-5	36	98	0 %
ssh-1986-l-side-2	36	154	13 %
hvg-2008-pages-l-1-2	34	83	33 %
meh-1967-l-side-1	27	101	0 %
hti-1977-l-pages-4-5	23	108	0 %
whe-1995-l-pages-9-10	22	48	52 %
h45-2000-l-side-1	19	54	33 %
hih-1984-l-pages-9-10	18	32	35 %
gno-2008-l-side-2	17	57	2 %
www-1999-l-pages-1-2	16	31	0 %
hdb-1969-l-side-2	15	26	35 %

the brochure *Helsinki Visitors Guide*, shown in Figure 8.8, (**hvg-2008-pages-l-1-2**) contains a high number of layout chunks and layout leafs: the entire layout structure of the double-page is visualised in Figure 8.9. I will now explain how the semiotic mode of page-flow operates at the intersection of layout and rhetorical structure using the example shown in Figure 8.9, and how its operation may be captured using the GeM model.

It is useful to begin with a few observations about the visualisation in Figure 8.9. The data were drawn from the layout layer using the XQuery script developed for visualising the layout structures and subsequently transformed into GraphViz DOT format (see Section 5.5.2). I then used OmniGraffle — a graphics editing software with a GraphViz engine<sup>1</sup> — to reorganise the hierarchical structure of the DOT graph into a radial form in order to make the visualisation fit on the page.

At the centre of the visualisation is the parent node which represents the double-page: the two nodes connected to the parent node stand for the individual pages. Below these two nodes, the layout structures of the individual pages exhibit a certain degree of similarity, but the structures are not symmetrical. A perfect symmetry would obviously require that the layout structures were identical. This kind of perfect symmetry is unlikely, given the unpredictable nature of multimodal semiosis on the more delicate levels, that is, the precise configuration of the structures identified in Chapter 7. In some other contexts, however, a perfect symmetry

---

<sup>1</sup><http://www.omnigraffle.com>

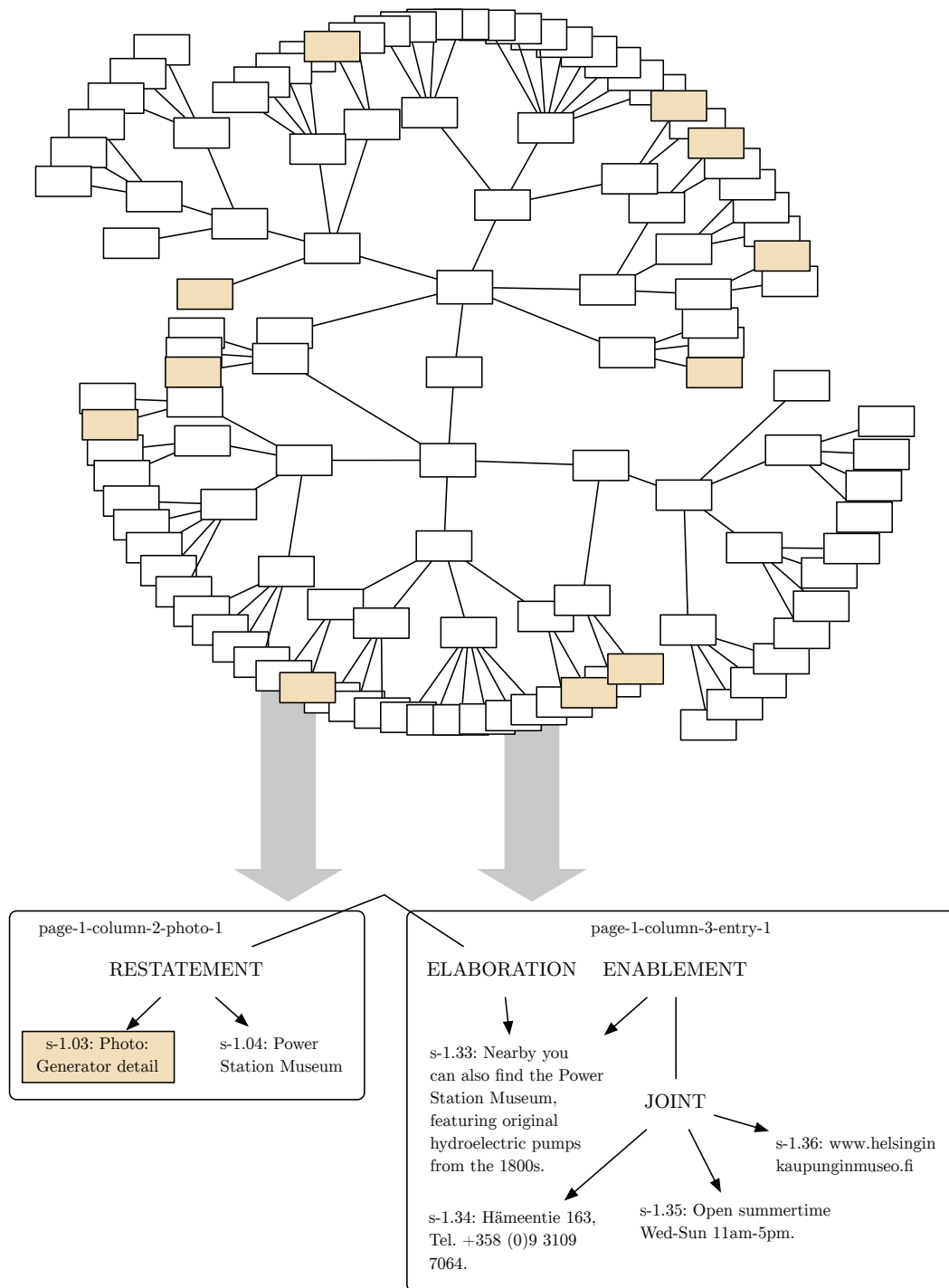


Figure 8.9: The layout structure and a part of the RST structure in *HVG 2008*

in the layout structure may serve a specific purpose, such as signalling equivalent meanings in bilingual content (cf. Berazhny and Hiippala in preparation).

Consequently, looking at the layout structure in Figure 8.9, the following question comes to mind: what enables the reader to put this kind of a complex asymmetric layout structure back together during the interpretation process? The high number of layout chunks ( $n = 34$ ) and layout leafs ( $n = 83$ ) appear to present a significant challenge, yet the readers seem to negotiate such structures with ease. I would argue that the successful interpretation of the complex layout structure is only possible because the structure of the layout chunks is not arbitrary. Instead, these structures are — as Bateman and Schmidt (2012, p. 48) point out — carefully articulated and functionally motivated, and for this reason, they often follow a consistent structural pattern. In short, these structures are “semiotically charged organisations” (Bateman 2011, p. 20).

These structures are exemplified in Figure 8.9, in which the lower half shows two common structures for realising the genre of a tourist brochure: a verbal description with an ENABLEMENT pattern (**page-1-column-3-entry-1**) and an image-text-complex (**page-1-column-2-photo-1**) (see Sections 7.1.3 and 7.2.1 respectively). Although these structures belong to different layout chunks in Figure 8.9, a rhetorical relation does hold between the RST segments in these layout chunks. To establish this relation, the reader is required to adopt an interpretation consistent with the semiotic mode of page-flow. In this case, the correct interpretation involves identifying a specific pattern: the visual representations of the verbal location descriptions are realised using image-text-complexes in marginal positions of the area model (see Figure 8.10). The captions are used to signal the rhetorical relation between text-flow and the image-text-complexes.

Considering that the pattern described above is repeated throughout the entire double-page, the view of document design as a “pattern language” (Waller et al. 2012) may well be an adept choice of describing multimodal meaning-making on a page (see also Section 4.4.3). With the analyses presented in this dissertation, we can already make headway into establishing the patterns relevant for the tourist brochures. Whether the configurations of the semiotic modes in the tourist brochures are referred to as articulated parts (Bateman and Schmidt 2012) or design patterns (Waller et al. 2012) is of secondary concern: the primary target of investigation should be how the semiotic modes facilitate the interpretation of the participating semiotic resources, both individually and together.

This task may be informed by two notions introduced in Chapter 4. Firstly, the notion of genre as an expectation-generating device — introduced in Section 4.4.1 — provides a starting point. According to Bateman, a genre creates expectations towards its content and communicative goals, and their realisation also anticipates specific “forms of expression” (2009b, p. 14). In the case of the tourist brochure as

a genre, these forms of expression correspond closely to those described in Chapter 7.

Secondly, these expressions may indeed be characterised as a pattern language (see Section 4.4.3), because they are *not* limited to the genre of a tourist brochure. To exemplify, a particular form of expression — such as the image-text-complex — may be found across other media and genres (cf. Knox 2007; Caple 2009*a*). For this reason, I would argue that the expectations towards both (1) content and (2) structure, and the ‘patterned’ response to these expectations facilitate the interpretation of the semiotic modes, which allows the readers to negotiate complex layout structures, such as that shown in Figure 8.9.

To conclude, I have argued above that the layout and rhetorical layers of the GeM model may be used to capture the multimodal structures responsible for generating expectations towards the content, its structure and their interpretation. For this process of interpretation, the capability to select the correct semiotic mode is at the heart of “document literacy” (Waller 2012, p. 241). And as recent studies in document recognition show, humans classify multimodal artefacts with a high degree of uniformity, which suggests that document literacy is sensitive to different artefact types (see e.g. Cohen and Snowden 2008; Chen et al. 2012). With these points in mind, the following section moves to discuss the discourse semantics, in order to establish how the interpretations provided by the various semiotic modes work with each other on a page.

### 8.3 Formalising the discourse semantics

Having already considered the semiotic modes from the perspectives of the layout and rhetorical structures, I now want to pay specific attention to one aspect of the semiotic modes which involves both of these structures, that is, their discourse semantic interpretation. In this section, I will argue that these structures play a part in signalling the reader how to interpret the page and its parts, therefore contributing directly to the discourse semantics of the deployed semiotic mode.

According to Bateman (2011, p. 33), a successful process of interpretation involves selecting and applying the discourse semantics of the correct semiotic mode to the multimodal artefact at hand. In this way, applying the correct discourse semantics allows the reader to access the content. Again, the current dissertation may be used to exemplify the process of interpretation. Because this dissertation uses text-flow to communicate its meanings, it is unlikely that anyone would attempt to interpret this page using the discourse semantics of *page-flow*, as nothing signals that additional meanings could arise from the use of layout space. In this case, the correct interpretation can be then achieved by applying the discourse

semantics appropriate for linguistic discourse in the genre of an academic dissertation.

In other cases, such as that of *HVG* 2008 in Figure 8.9, the application of the discourse semantics of page-flow is far more likely. I base my argument on the presence of multiple semiotic modes on the page, which simultaneously introduce several possible discourse semantic interpretations. What the reader has to do, then, is to establish appropriate interpretations for each semiotic mode. Furthermore, Bateman (2011, p. 34) has suggested that the discourse semantic interpretations may also compete with each other on the page. However, before it becomes possible to evaluate whether the interpretations are in competition with each other, the discourse semantics need to be set out clearly for each semiotic mode deployed on a page.

To do so, let us continue with the current example in *HVG* 2008. Figure 8.10 shows the area model of a single page in *HVG* 2008, in which the baseline grid has been marked for its contents. In total, there are four image-text-complexes for the visual representations and three instances of text-flow, which realise verbal location descriptions. Additionally, a header and a small map occupy their own layout areas. To establish how this page is to be interpreted, I will now take the required discourse semantic interpretations apart.

Beginning with the semiotic mode of text-flow, it should be kept in mind that text-flow is characterised by its linear structure. In Figure 8.10, text-flow is used to realise several stages of the genre of a tourist brochure, which were set out in Section 4.3.1. As the analysis in Section 7.1 showed, the rhetorical layer of the GeM model may be used to capture the discourse structure of text-flow, and how it structures different stages of the genre, such as location descriptions or travel and contact information. This is exemplified by the *ENABLEMENT* pattern shown in Figure 8.9, which provides the reader with travel and contact information.

The interpretation of text-flow is naturally influenced by the fundamental principle of linearity, which lies at the core of this semiotic mode. This is represented in Figure 8.11a, which illustrates the discourse semantic interpretation of text-flow using a back-and-forth mapping (Bateman 2011, p. 28). Note, however, that Figure 8.11a is an abstraction that shows the principle behind text-flow. In text-flow, the unfolding discourse is subject to order and text structuring on various strata of language (see e.g. Martin 1992, 1994). It is therefore important to distinguish Figure 8.11a from the detailed descriptions of text-flow achieved using RST in this dissertation. Although the rhetorical structures are governed by the principle of linearity, they constitute a more detailed representation of the discourse structure of text-flow, which is valid specifically for the genre of the tourist brochure.

This raises the following question: how does one interpret text-flow when it participates in the semiotic mode of page-flow? Recall that Figure 8.10 contains

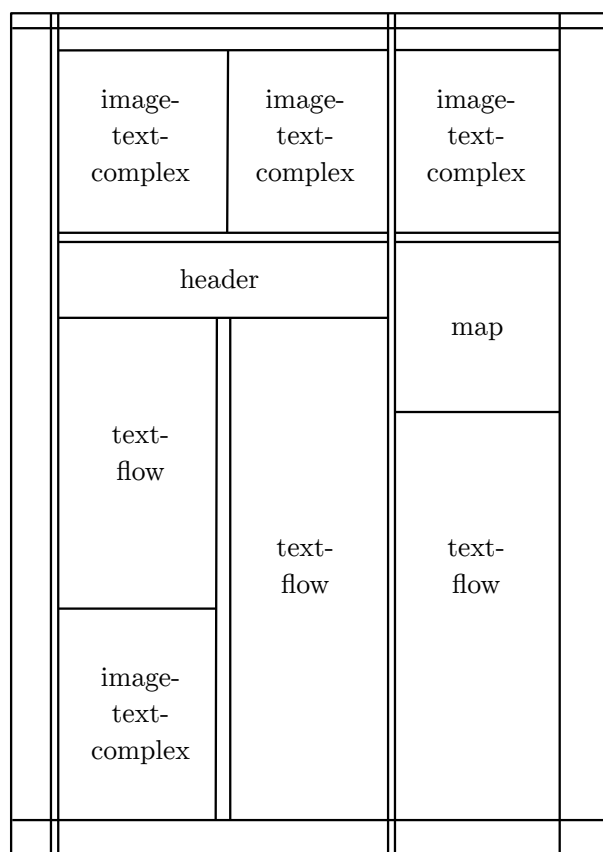


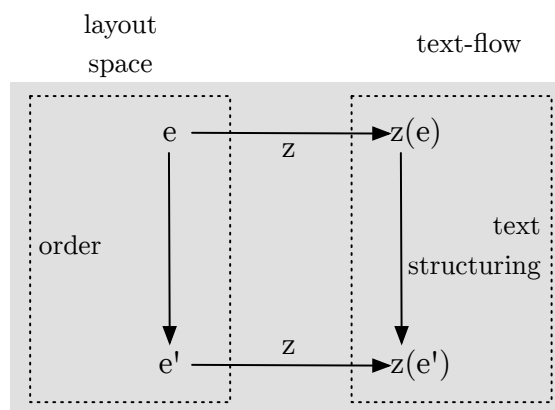
Figure 8.10: Area-model of a single page in *HVG* 2008

three instances of text-flow. First of all, ‘zooming out’ of text-flow and moving to observe the entire page layout requires abandoning the principle of linearity (cf. Waller 2012, p. 241). This is reflected in the back-and-forth mapping in Figure 8.11b, which sets out the principle behind the discourse semantic interpretation of page-flow. The general principle behind page-flow is elaborated by Bateman (2011, pp. 28-29):

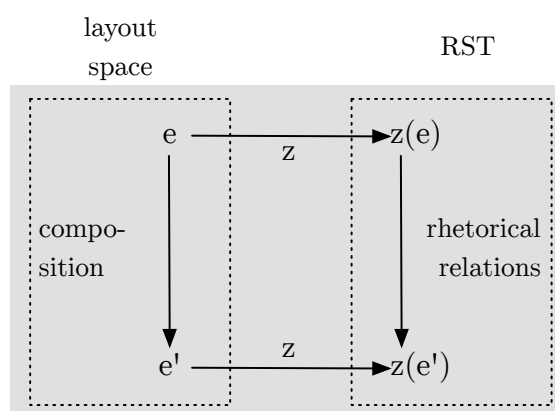
[In page-flow] we find composition relations (i.e. spatial layout brought together with the possibilities of rhetorical structure theory: in short, the hierarchical composition is placed in correspondence with hierarchical rhetorical organisation.

In GeM terms, the “hierarchical composition” corresponds to the layout structure and the area model. As Figure 8.9 showed, rhetorical relations may hold between different layout chunks in the hierarchical layout structure. Furthermore, these layout chunks may be physically positioned in different layout areas (see Figure

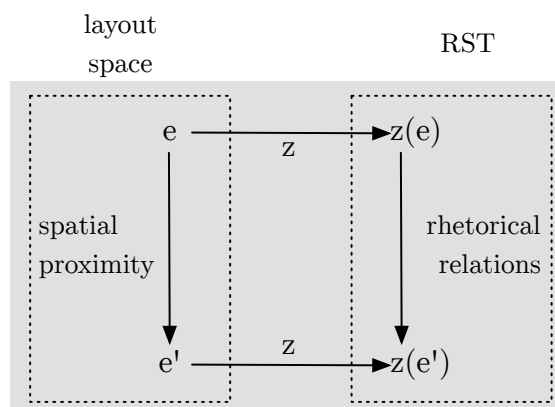




(a) Text-flow (after Bateman 2011, p. 28)



(b) Page-flow (after Bateman 2011, p. 28)



(c) Image-text-complex

Figure 8.11: Back-and-forth mappings of discourse semantics in *HVG* 2008

8.10). As such, *HVG* 2008 is a fine example of page-flow in action. By focusing on the interface between the layout and rhetorical structure, the specific configuration of page-flow in this brochure becomes available for analytical description using the GeM model. This naturally anticipates the next question: what triggers the discourse semantic interpretation associated with page-flow in the case of *HVG* 2008?

One possible answer may be suggested by contrasting the layout structures illustrated in Figures 8.7 and 8.9, which represent two different semiotic modes. The first example in Figure 8.7 represents a layout structure consistent with text-flow. The structure is characterised by a low amount of layout chunks and leafs, and a shallow layout structure. Due to the simple layout structure, I would argue that putting the structure of this page back together during the interpretation process involves less work than in the case of Figure 8.9. This results from the fact that the layout structure in Figure 8.9 is more complex in several aspects. Firstly, the double-page in Figure 8.9 has a higher number of layout chunks and leafs, and secondly, its layout structure is deeper and fragmented into multiple chunks and leafs.

In the light of these observations, I suggest that the layout structure plays a significant role in triggering an interpretation based on the discourse semantics of page-flow. The presence of multiple layout chunks triggers an interpretation that encourages the reader to check whether rhetorical relations hold between the layout chunks. The identification and interpretation of these layout chunks is in turn facilitated by particular forms of expression used in the tourist brochures. These relevant forms of expression identified in the annotated corpus were described in detail in Chapter 7 by combining the information from the layout and the rhetorical layers of the GeM model.

With the generic principles behind page-flow now established, the next step is to achieve a more accurate description of how this semiotic mode operates in the tourist brochures. The previous descriptions that rely on the layout and rhetorical layers may be complemented by the back-and-forth mappings used to represent the discourse semantic interpretation of the semiotic modes. Consider, for instance, the image-text-complexes in Figure 8.10. I previously argued that the meaning potential of the image-text-complex arises from the spatial proximity of the image and the accompanying caption (see Section 7.2.1). The discourse semantic interpretation that guides the interpretation of the image-text-complex is represented using a back-and-forth mapping in Figure 8.11c.

Note that for the image-text-complex in Figure 8.11c, the relation holding between the entities  $e$  and  $e'$  in the domain of layout space is given a specific definition. The spatial proximity of the entities in the layout space indicates that a rhetorical relation holds between them. Obviously, the scope of this interpretation

is limited in comparison to the abstract representations given for text-flow and page-flow in Figures 8.11a and 8.11b: the interpretation may only be applied to the image-text-complexes. Moreover, it needs to be understood that the image-text-complex is not a semiotic mode, but an established multimodal structure — a form of expression — which may appear as a part of both text-flow and page-flow (cf. Bateman 2009*b*). To sum up, the back-and-forth mappings may complement the descriptions provided by the GeM model, particularly in the case of complex multimodal structures.

Finally, I wish to emphasise that the semiotic modes are abstractions, which provide an efficient analytic tool for describing meaning-making in multimodal artefacts. The discourse semantic definitions given in Bateman (2011, p. 28) establish the operating principles of the semiotic modes, but additional work is required to expose how exactly these semiotic modes are configured in specific contexts, that is, in other multimodal artefacts. An example may be drawn from the study of filmic discourse which is, in principle, a form of dynamic image-flow. The work of Bateman (2007), Bateman and Schmidt (2012) and Tseng and Bateman (2010, 2012) has shown that the syntagmatic and paradigmatic choices in the semiotic mode of film can be mapped out to a considerable extent, while also explaining the impact of these choices on the structure and interpretation of discourse in different filmic artefacts.

In many ways, the current position of this dissertation resembles that of the filmic analyses described above. Having already established the properties of the medium used — the brochure — in Chapter 6 and the semiotic modes that participate in this medium, the various structural configurations analysed in Chapter 7 provided an in-depth understanding of the semiotic modes and their specific configurations. These configurations, which may be described as patterns or established means of expression, realise the genre of the tourist brochure using the medium of a brochure, and eventually come together to form a multimodal artefact. Finally, with the semiotic modes and their features identified in this chapter, the dissertation now stands in a good stead for the final task: investigating the distribution of the semiotic modes in the annotated corpus.

## 8.4 Text-flow and page-flow in the annotated corpus

In this section, I investigate the distribution of the semiotic modes in the annotated corpus by building on the configurations identified in Chapter 7 and the previous analyses of the layout structure in Section 8.2.1. This work is used to define the criteria for querying the annotated corpus for particular configurations of the

semiotic modes. With these criteria, it becomes possible to track the use of the semiotic modes over time. This is of great interest to a model of artefact structure, because it provides a vantage point for observing the impact of the semogenic processes on the artefact structure (see Section 4.3.2.1).

### 8.4.1 Detecting the active semiotic mode

Let us begin with Figure 8.12, which presents three types of data related to the semiotic modes and their use. Firstly, the grey and blue bars in the background indicate the percentage of layout space occupied by graphic elements: I will explain the use of colour shortly below. The percentages for layout space are given on the vertical axis on the right-hand side. The vertical axis on the left-hand side indicates the number of items for two types of data: image-text-complexes (solid line) and text-flow paragraphs (dashed line).

The three data sets shown in Figure 8.12 were retrieved from the annotated corpus using three different queries, whose criteria I will now set out in greater detail. Note that the procedure of retrieving the data on the use of layout space, represented by the grey and blue bars, was already described in Section 6.3.1. For all queries, the data were retrieved from the segmentation, layout structure and realisation information components of the layout layer. With the basics of Figure 8.12 now explained, I shall move to discuss the image-text-complexes in the annotated corpus.

I decided to focus on the image-text-complexes that combine a photograph with a verbal description due to the prominence of photographs in the annotated corpus (see Table 7.2). I proposed earlier that this structure may be used to integrate visual and verbal content for providing specific visual representations in the tourist brochures (see Section 7.2.1). I also suggested that image-text-complexes use the layout space to establish a relation between the participating entities. The correct inference of this relation, in turn, requires applying a specific discourse semantic interpretation (see Figure 8.11c).

For identifying the image-text-complexes in the annotated corpus, I defined the following criteria:

1. The layout chunk consists of two or more layout leaves.
2. One layout leaf is a layout unit of the type *photo*.
3. One layout leaf is a layout unit consisting of one or more verbal base units.

With these criteria, the query identified a total of 71 image-text-complexes. Due to the several types of image-text-complexes identified in the tourist brochures — such as captioned photographs and illustrated descriptions — I specified that a verbal layout leaf could consist of one or more verbal base units, because this would accommodate both of the aforementioned structures.

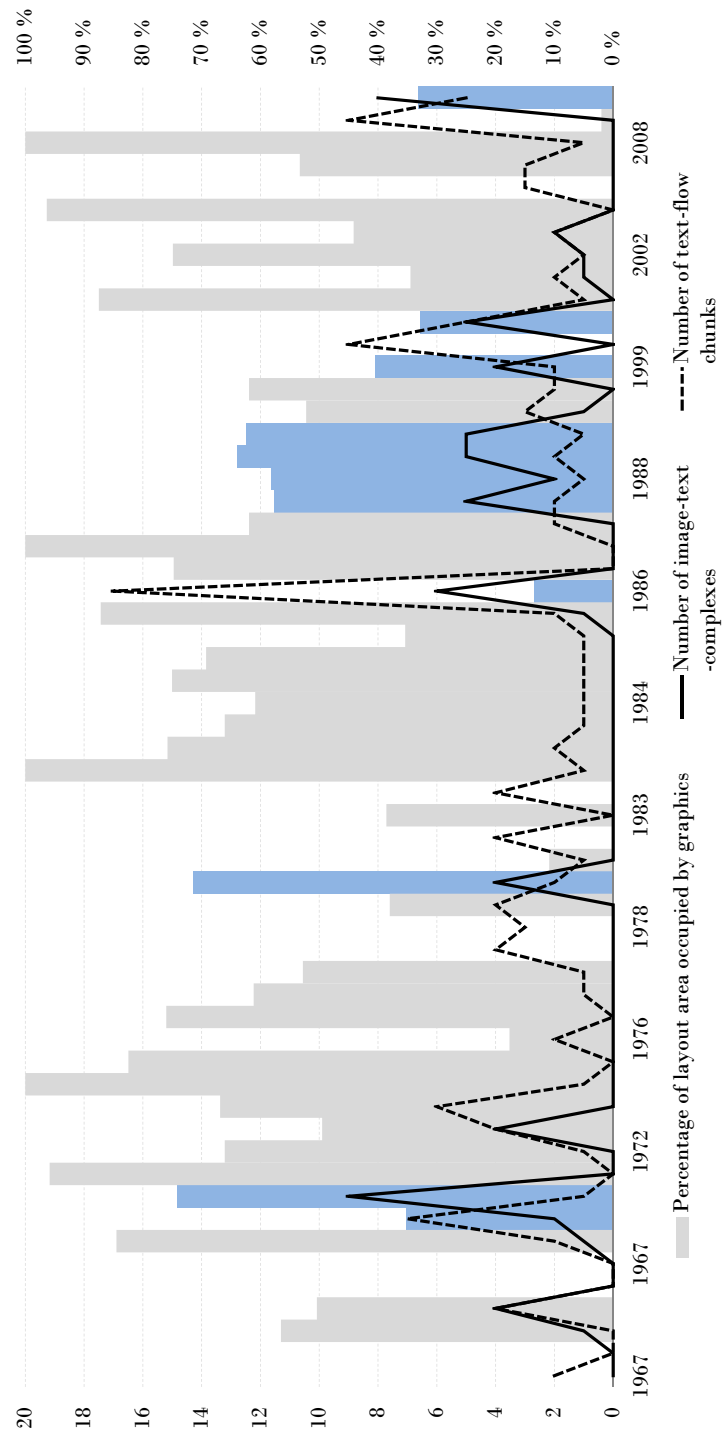


Figure 8.12: The distribution of semiotic modes in the annotated corpus

The next step is to consider the distribution of image-text-complexes in the data. As Figure 8.12 shows, the annotated corpus includes many pages without any image-text-complexes, especially in the late 1970s and early 1980s. The image-text-complexes seem to appear sporadically, except during the late 1980s. It should be noted, however, that this period includes the pages of a single brochure (*HFT* 1988), which was fully annotated during the pilot project (Hiippala 2012*b*).

What may be observed in Figure 8.12 is that the image-text-complexes often accompany each other. In 14 out of 20 cases, a page with an image-text-complex includes more than just one image-text-complex. My hypothesis is that the presence of multiple image-text-complexes may indicate that the layout structure of the page is likely to be fragmented and thus susceptible to a discourse semantic interpretation based on page-flow (cf. Figures 8.5 and 8.9). This is exemplified by *SSH* 1986, which stands out in Figure 8.12 with its high number of layout chunks with text-flow. I shall elaborate my hypothesis below, after considering the distribution of text-flow.

I defined the following criteria for identifying instances of text-flow in the annotated corpus:

1. The layout chunk consists of two or more layout leafs.
2. All layout leafs are of the type *text*.
3. The layout leafs consist of three or more verbal base units.

To distinguish the instances of paragraphed text-flow from headers, captions and other short verbal elements, I set additional constraints for text-flow. These constraints are explained below.

Within the layout layer of the GeM model, the most common linguistic unit is typically a paragraph (Bateman 2008, p. 116), which often consists of multiple sentences. To exclude short paragraphs of two sentences or less, I set the criteria to include only those paragraphs with three or more verbal base units. This would exclude short text fragments in the data. I assume that these criteria were sufficient for identifying most cases of paragraphed text-flow in the data, as the query returned a total of 137 instances.

As Figure 8.12 shows, certain pages do not include any instances of text-flow. These pages fall into two categories. The first category consists of the pages without visual elements, as indicated by absence of the grey and blue bars (cf. 1967). These pages are either table- or list-driven (see Section 7.1.1). The second category is that of cover pages, which do not include instances of text-flow, but the presence of visual elements may be detected (cf. 1976). The pages that do not belong to either of these categories may be then considered from the perspective of the semiotic modes: can we use the data presented in Figure 8.12 to infer the semiotic mode active on the page?

I now present two criteria for identifying the pages that deploy page-flow in the data. Firstly, it should be noted that image-text-complexes may also occur within text-flow. The image-text-complexes or other visual structures may either interrupt the paragraphed text or they may be placed alongside text-flow (see e.g. Royce 1998; Matthiessen 2007). In contrast, I consider the presence of *two or more image-text-complexes on the same page* to signal the use of page-flow: this constitutes the first criterion.

The second criterion is that the *number of image-text-complexes and text-flow chunks does not match*. As the layout structure in Figure 8.7 showed, the layout structure of text-flow can be relatively shallow. For this reason, an instance of text-flow and an accompanying photograph may be positioned under the same layout chunk. And because no upper limit was defined for the number of verbal base units participating in the image-text-complexes, these instances of text-flow are identified by the query as image-text-complexes (cf. 1967, 1972, 2002). In many cases, these pages actually consist of paragraphed text-flow and an accompanying conceptual photograph (see Sections 7.1 and 7.2.2 respectively).

#### 8.4.2 A closer look at page-flow

In Figure 8.12, the blue colour marks the instances of page-flow, which were identified either using the criteria presented in Section 8.4.1 or by analysing the rhetorical structure as described in Section 8.1. What characterises many instances of page-flow is the simultaneous occurrence of multiple image-text-complexes and text-flow chunks, which suggests that the content is fragmented (cf. Figures 8.5 and 8.9). The ratio between image-text-complexes and text-flow chunks is likely to differ, because these forms of expression are made available by the genre of a tourist brochure.

These forms of expression are deployed according to the communicative needs of the artefact. For some purposes, text-flow may be the most appropriate choice — for example — to realise multiple location descriptions during a “stop-look-see” walking tour of the destination (cf. Enkvist 1991, p. 9). Page-flow, in contrast, allows the page to break away from the linear structure of language (Waller 2012, p. 239), and this may encourage the use of visual configurations — such as image-text-complexes — to realise various stages of the genre. This proposition is supported by the number of pages identified by the query using the layout structure.

However, even if the layout structure may provide many cues for identifying the deployed semiotic mode, the input of the rhetorical layer may be required in certain cases. One such case is *WEH* 1998, a brochure discussed in Section 8.1, in which the layout structure resembles that of visually-enriched text-flow. The presence of page-flow could only be identified by looking at the rhetorical structure, which supports the bid that cross-layer analyses are required to tease

out the principles behind the structure of an artefact. This is exactly where the strength of the GeM model lies, although there has been certain misunderstanding about the model's multiple analytical layers and their contribution to the analysis of multimodal artefacts (cf. Scott 2010).

In Section 3.5, I argued that understanding how page-flow operates is important for a model of artefact structure. On the basis of the annotated corpus, I would now argue with relative confidence that when page-flow is used to realise the genre of a tourist brochure, its main function is to segment and organise the content into easily digestible units. In the tourist brochures, page-flow breaks the content into the recognisable structures described in Chapter 7. Compared to certain educational genres (see e.g. Bateman 2011; Hiippala 2012*c*), the functions and the resulting structure of page-flow are relatively uniform in the tourist brochures. Yet the annotated corpus showed that page-flow may also take more complex forms which exploit the two-dimensional layout space (cf. Section 8.1).

Before the conclusion, a final issue related to text-flow and page-flow needs to be considered, that is, their distribution in the annotated corpus. The capability to distinguish between text-flow and image-flow presents us with an opportunity to consider whether technological development, and particularly the emergence of desktop publishing have affected the deployment of the semiotic modes. This issue is taken up in the following section.

## 8.5 Semiotic modes and semogenic processes

In Section 6.3.2, I discussed the impact of technological development on the *medium* of a brochure, emphasising how the production and design processes are no longer dependent on the material substrate of the printed page. Because the designed artefacts do not have to be realised physically, I asked whether these changes have fostered experimentation in the design process. Consequently, the next question that needs to be asked is whether the technological development has affected the semiotic modes?

In contrast to the material substrate, the semiotic modes may be considered to be immaterial. They provide the means to realise meanings using multiple semiotic resources, but these meanings take a concrete form only when they are selected for participation in a medium. For the genre of a tourist brochure, the dissertation now possesses a substantial amount of knowledge about the brochure as a medium *and* the selection and configuration of the semiotic modes. With this knowledge, we may proceed to investigate whether technological development has indeed affected the choice of the semiotic mode.

To do so, let us consider the year 1985 as a turning point for desktop publishing. In 1985, Adobe introduced PageMaker, which quickly emerged as the industry



Table 8.3: Page-flow and text-flow in the annotated corpus

	Before 1985	After 1985
Text-flow	31	15
Page-flow	3	9
P-value	<0.05	

standard for DTP software. The tourist brochures may thus be divided into two categories: those published before ( $n = 34$ ) and those published after 1985 ( $n = 24$ ). With the distribution of text-flow and page-flow for the artefact available, we may divide the data into categories according to the deployed semiotic mode. Table 8.3 shows the data in a contingency table, in which its distribution may be studied using a two-tailed Fisher's exact test. As Table 8.3 shows, the difference in the distribution of text-flow and page-flow before and after 1985 is statistically significant.

It appears that the introduction of DTP may have affected the selection of semiotic modes and consequently, the structure of the tourist brochure as a multimodal artefact. At this point, however, several issues about the statistical analysis have to be raised. On the one hand, the modest size of the annotated corpus needs to be acknowledged ( $n = 58$ ). On the other hand, the annotated corpus was intended to represent the *series* of brochures that made up the entire data set. It is likely that only one semiotic mode can be dominant within an artefact, which lends credibility to the analysis and the representativeness of the annotated corpus. I would thus argue that these observations should be taken into serious consideration in future work on the 'visual turn' (see Section 6.3.1).

The above observation that the structure of the tourist brochures has shifted towards page-flow is particularly interesting in the light of another contingency table. In Section 6.3.1, I established that in the annotated corpus, an average of 50.5% of the layout space is occupied by graphic elements. If we take the previous division of data into the categories of pre- and post-1985, and use the 50% mark as a watershed between two additional categories, we can also consider the impact of desktop publishing on the degree of visuality in the brochures, as shown in Table 8.4.

As Table 8.4 shows, the application of a two-tailed Fisher's exact test indicates that there is no statistically significant difference in the degree of visuality before and after the year 1985. Thus it appears that the tourist brochures have been fairly visual throughout the time period represented in the annotated corpus. On the basis of this observation, I am tempted to propose that the impact of technological development has affected the multimodal *structure* of the brochures and their

Table 8.4: The degree of visuality in the annotated corpus

	Before 1985	After 1985
Graphics <50%	14	9
Graphics >50%	20	15
P-value	0.79	

internal logic, while their visual appearance and the proportions of language and image have remained the same.

As Molina and Esteban (2006) have established, visuality constitutes an important factor for establishing a positive view of the advertised destination in the reader's mind. Therefore, it is not surprising that the tourist brochures continue to be highly visual, because it has been established as an efficient means to influence the tourists (see also Scarles 2004; Jokela 2011). Yet the degree of visuality has little to do with an artefact's multimodal *structure*, that is, how the graphic elements are integrated into the structure of the tourist brochures. This clearly anticipates two avenues for further work: (1) the analysis of the content represented in the visuals, and (2) their participation in the structure of a multimodal artefact, to which this dissertation has contributed.

Finally, one more aspect of the semiotic modes of text-flow and page-flow may be considered, that is, their distribution across the brochures and leaflets (see Section 6.1.2). If the brochures and leaflets indeed constitute two *medial variants* within the medium of a brochure, the semiotic modes are potentially a domain where the two medial variants differ. To pursue this line of investigation, the data in the annotated corpus were grouped into a contingency table according to the semiotic mode and the medial variant, as shown in Table 8.5.

Table 8.5: The semiotic modes in the brochures and leaflets

	Brochures	Leaflets
Text-flow	26	20
Page-flow	8	4
P-value	0.74	

According to Table 8.5, the difference in the distribution of the semiotic modes across the medial variants of brochures and leaflets is not statistically significant. This may suggest that the medial variants arise during production as a result of the need to include more content pages into the artefact (cf. Figure 6.1). In these

medial variants, both text-flow and page-flow may then be used to realise the genre of a tourist brochure.

This concludes the analysis presented in this chapter. Next, before the discussion of the dissertation and its results, I will present a brief summary of this chapter.

## 8.6 Summary

The aim of this chapter was to analyse the page as a part of the tourist brochures. I approached this task from the perspective of the semiotic modes, drawing on the knowledge of the multimodal structures acquired in Chapter 7.

I began by pointing out in Section 8.1 that the semiotic mode active on a page cannot be determined according to the degree of visuality or any other superficial characteristic of the page. In contrast, within the GeM framework, determining the active semiotic mode requires inputs from both layout and rhetorical layers. This provided the basis for a stronger, structure-based argument for determining the active semiotic mode.

In Section 8.2, I directed the analytical attention towards the semiotic mode of page-flow. By taking apart the layout structure, I established several differences between page-flow and text-flow. The layout structure of page-flow is more complex in terms of the number of layout chunks and leafs, which are often organised into a deeper hierarchy. I proposed that in order for the reader to be able to make sense of these complex layout structures, their organisation needs to be functionally motivated and carefully articulated. Moreover, the organisation needs to follow the typical forms of expression used in the tourist brochures, which I described in Chapter 7.

The process of interpreting a page was considered in Section 8.3 from the perspective of discourse semantics. I argued that although the semiotic modes are abstractions, they can be used to capture the principles behind different semiotic modes. A closer analysis, however, is required to tease out their actual configurations, to which this dissertation has contributed.

Having outlined the generic principles behind text-flow and page-flow, I set out to examine how they may be identified in the annotated corpus in Section 8.4. In most cases, the layout structure provided sufficient information to identify the semiotic mode active on the page. In some cases, however, additional input from the rhetorical layer was required.

Finally, with the semiotic modes identified, I proceeded to observe their distribution in the annotated corpus. Having considered the impact of technology in Section 6.3.2, I set out to investigate whether the introduction of desktop publishing had affected the deployment of the semiotic modes. In this case, a statistically

significant difference was found in the distribution of text-flow and page-flow before and after the year 1985, which marked the introduction of desktop publishing. In contrast, differences in the degree of visuality or the choice of the medial variant — brochure or leaflet — were not found to be statistically significant.

This section concludes the analyses undertaken in this dissertation. I will now move to discuss the results of the study in Chapters 9 and 10.

# Chapter 9

## Discussion

### 9.1 Main findings

The goal of this dissertation was to model the structure of the tourist brochure as a multimodal artefact. To pursue this goal, the dissertation set out to answer three research questions, which were presented in Section 1.5.1. In the following sections, I will evaluate each research question in the light of the performed analyses and discuss how the findings contribute towards the goals of this dissertation.

#### 9.1.1 Which factors affect the structure of a multimodal artefact?

The results corroborated the previous research on print media, which has suggested that the material substrate and its manipulation during production affect the multimodal structure of the artefact (see e.g. Kress and van Leeuwen 2001; Bateman 2008, forthcoming). In this dissertation, I addressed the established forms of material substrate using the notion of medium, which is itself a theoretically loaded concept with many conflicting definitions.

However, by drawing on the two data sets, I was able to determine more precisely what the medium contributes to the multimodal artefact. I approached the issue in Chapter 6, focusing on the material substrate at the heart of the medium: the printed page, its method of binding and fold geometry.

Using these two features, I identified two *medial variants* within the medium of a brochure: staple-bound brochures and folded leaflets. The staple-bound brochures increased their semiotic space by adding pages to the artefact, while the leaflets expanded their semiotic space by introducing additional folding points. The available semiotic space, determined by the amount of content pages, was significantly higher in the staple-bound brochures. The method of binding using

staples appeared to be more efficient for increasing the semiotic space, without constraining the use and physical handling of the artefact.

However, the choice of a medial variant did not influence the choice of semiotic modes in an artefact. After the analysis of multimodal structure, which provided the means to detect the active semiotic mode, I explored the possibility that the staple-bound brochures and the folded leaflets would differ in their use of the semiotic modes (cf. Valdeón 2009). The results showed that both medial variants were free to deploy both text-flow and page-flow.

Following Bateman (forthcoming), I also outlined a set of features that should be considered to arise from the medium. These features, such as page numbering, text spacing and margins, structure the artefact and assist the reader. All artefacts realised using the medium of a brochure may include these features and for this reason, their contribution has to be clearly separated from the content.

The same applies to advertising, which can populate a medium and contribute semiotic material to the artefacts in the form of classified and full-page advertisements. Yet advertisements were relatively rare in the data, with classified advertisements appearing only after the year 2000 in staple-bound brochures with a high number of content pages. The advertisements were also used to extend the overall functions of the artefact, as certain brochures could be used to redeem discounts at the advertised locations.

To sum up, the results suggested that the medium determines the physical size and form of the artefact, but the choice of the medial variant did not influence the use or configuration of the semiotic modes. Moreover, the medium also contributed semiotic material to the artefact in the form of advertisements and navigation structures. In modelling the structure of an artefact, these features needed to be considered independent of the content, whose structuring I approached with the next research question.

### **9.1.2 What kinds of patterns may be identified in the multimodal structure?**

Following the recent advances, which have postulated that multimodal structures are patterned and functionally motivated (see e.g. Holsanova and Nord 2010; Bateman and Schmidt 2012; Waller et al. 2012), I chose the annotated corpus as the point of departure for searching patterns in the multimodal structure of the tourist brochures. Rather than imposing predefined patterns on the data, I drew on the work of Thomas (2009a) in developing a visualisation technique, which combined information from the multiple analytical layers of the GeM-annotated corpus to assist in the search for patterns.

I thus focused on two types of structures: the hierarchical organisation of the content in the layout structure and its presentation to the reader in the rhetorical structure. By observing the annotated corpus from these perspectives, I was able to establish how the functionally motivated patterns are construed in the tourist brochures. This approach proved fruitful and helped to identify several patterns in the data. From a theoretical perspective, these patterns may be considered the “significant clusters of variables within and across modes” (Forceville 2007, p. 1237) and the familiar “forms of expression” (Bateman 2009a, p. 14) used to realise the genre of a tourist brochure.

I described these patterns as particular configurations of the deployed semiotic modes, which were responsible for various communicative tasks. For instance, specific configurations of text-flow were used to provide the reader with descriptions, travel and contact information, and guidance. In addition, I identified two configurations used for integrating graphic elements into the multimodal structure of the brochures: image-text-complexes (cf. Kvåle 2010) and conceptual photographs. To sum up, the multimodal structure provides the skeleton of the genre, which is then fleshed out by the content. Together, they create expectations towards the multimodal artefact in the mind of the reader. This feedback loop between the reader and the designer, in turn, may be essential for sustaining the genre.

Finally, I was able to connect the patterns observed in the data to the “principles behind the structure” (Holsanova and Nord 2010, p. 83), which I described using the notion of semiotic modes (Bateman 2009b, 2011). Obviously, the semiotic modes are high-level abstractions, but with the help of the annotated corpus, I established a connection between the observed multimodal structures and the semiotic modes. In the artefacts, text-flow was mainly used for extensive descriptions, while page-flow segmented these descriptions into easily digestible units. Their structural difference emerged particularly in connection with the layout structure. This observation, in turn, proved highly relevant to the next research question, which was related to the processes of change.

### **9.1.3 How does the structure of the multimodal artefacts change over time?**

By building on the model of change presented in Figure 3.1 and the structural analyses of the semiotic modes described above, I was able to tease out the specific configurations of two semiotic modes active in the tourist brochures: text-flow and page-flow. This provided the means to detect the semiotic mode active on the page and observe their occurrence in the annotated corpus between the years 1967 and 2008, which provided an empirical perspective to the so-called ‘visual turn’ in communication.

A major finding was the statistically significant difference between the distribution of text-flow and page-flow before and after the year 1985. This year marked the introduction of desktop publishing, which appears to have affected the use of the semiotic modes in the tourist brochures. Whereas the artefacts produced before 1985 mainly use text-flow, the semiotic mode of page-flow became more prominent after 1985. This suggests that the underlying structural principles of the page have changed towards an organisation which exploits the entire two-dimensional space on the page.

Despite the fundamental changes in the processes of design and production resulting from the introduction of desktop publishing, certain aspects of the tourist brochures have remained unchanged. For instance, an average of 50.5% of the layout space on the pages of the annotated corpus was occupied by graphic elements. However, there was no statistically significant difference in the use of layout space before and after 1985. It appears that the tourist brochures have been visual throughout the period, and the semogenic processes have mainly affected their multimodal structure.

To conclude, the annotated corpus provided valuable insights into the multimodal structure of the tourist brochures. What made the longitudinal observations possible was the detailed and systematic description of artefact structure in the corpus, whose data could be queried and visualised according to need. Moreover, the corpus-driven analyses presented in this dissertation also set a new standard for critical studies of the ‘visual turn’, which Kostelnick and Hassett (2003, pp. 230-231) have called for in the field of information design. Next, I will focus on the relevance of the findings presented above.

## 9.2 Relevance of the findings

Keeping the principles of theory-building in mind, I will now consider the relevance of these findings to the field of multimodal research. The question is: what can be fed back into the theory of multimodality on the basis of the findings?

To begin with, modelling the structure of a multimodal artefact provides the much-needed middle ground between the detailed analyses of multimodal meaning-making (Royce 1998, 2007; Kress and van Leeuwen 2006; O’Halloran 2008*b*) and the more abstract theoretical considerations (see e.g. Lemke 2000, 2005; Kress and van Leeuwen 2001). Up to this point, this middle ground has been an area where multimodal research has not measured up to the expectations. For instance, Royce (1998, p. 46) acknowledged at an early stage that the studies of multimodal cohesion would benefit from a theory of multimodal structure. I would argue that this need results from the structural complexity of multimodal phenomena, which Lemke (2000, p. 100) has described as follows:



Certainly for biological systems, and probably for many others as well, the richness of their complexity derives in part from a strategy that organises smaller units into larger ones, and these in turn into still larger units, and so on.

The model of artefact structure accounts for this complexity by identifying the various sources and levels of organisation in multimodal meaning. To use a spatial metaphor, in the horizontal plane we may use the notions of medium, semiotic modes and genre to model artefact structure, which enables us — once a model is established — to move in the vertical direction. In this direction, we may cast our attention above to the relations between genres and their structural similarities and dissimilarities (see e.g. Waller et al. 2012; Hiippala in press), or alternatively, below to the domain of semiotic phenomena described in the detailed analyses: cohesion, metaphor, intersemiotic and image-text relations (Forceville 1996; O’Halloran 1999*a*; Martinec and Salway 2005; Martinec 2013).

We can now better situate the studied phenomena, which may have profound implications to the theory of multimodality. As Hasan (2004, p. 16) has argued, it is imperative for the process of theory-building to establish the relations between the used theoretical concepts. One such example identified in this dissertation is the difference between the multimodal notions of coherence (Bateman 2008, p. 151) and cohesion (Royce 1998, 2007), which I will now describe to illustrate my point.

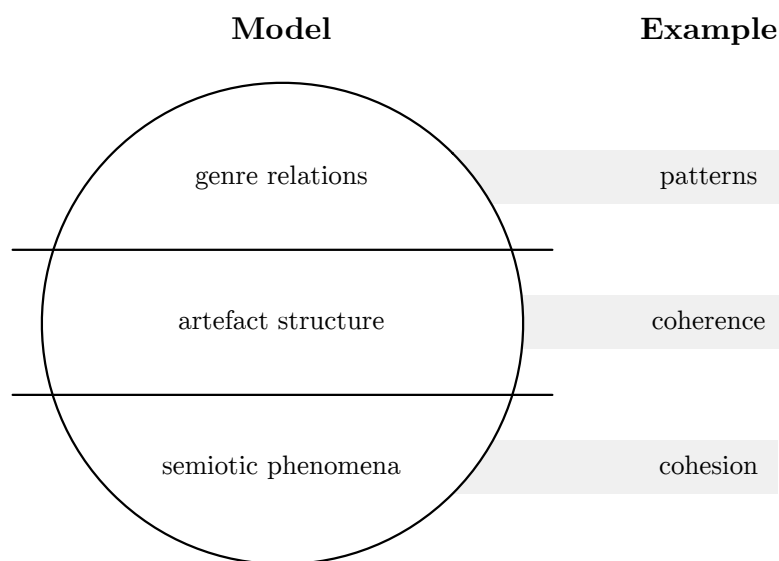


Figure 9.1: Situating the model of a multimodal artefact

Figure 9.1 shows how coherence and cohesion may be situated within a wider framework. I would propose that coherence is a property of multimodal structure,

which is concerned with the artefact and its commitment to the communicative goals. Whether the artefact achieves these goals depends on how well it meets the readers' expectations towards its multimodal structure. As I have showed in this dissertation, the artefact structure arises from the combination of the used medium, the deployed semiotic modes and the genre that is to be realised. In this way, coherence is directly concerned with the structure of a multimodal artefact.

Cohesion, in turn, should be understood as a semiotic phenomenon that arises from the deployment of multiple semiotic modes to communicate the same subject matter. This phenomenon is the product of very delicate choices in the semiotic resources, and it is thus best modelled using the frameworks with an extensive analytical reach, such as those described in Chapter 2. However, to remove these observations from their imminent context for a broader perspective, the analysis needs to be supported by a notion of multimodal structure, such as that of coherence. In short, to respond to the analytical challenges presented by the wealth of multimodal phenomena around us, we need proven theoretical concepts and these concepts need to be well-formulated and connected to each other.

Finally, I want to raise the issue of change. As Kress (1998, p. 74) has argued, a well-formulated theory of semiosis needs to account for change. In modelling the structure of a multimodal artefact, I tackled this issue by using the notion of semogenic processes. By building on the developed model of artefact structure and the annotated corpus, I showed that the changes in the tourist brochures have been structural as opposed to merely visual.

A plausible explanation may be provided by considering the multimodal structure and the realisation of an artefact. As I showed, the tourist brochures have been visual for a long time already, but their internal structure has shifted towards page-flow following the introduction of desktop publishing. At the same time, the visual appearance of the brochures has changed frequently. These rapid changes in the graphic and typographic realisation are possible, because they do not require new models of interpretation in terms of the semiotic modes. For the same reason, the artefacts do not allow rapid changes to their internal structure, because the configurations of the semiotic modes for realising a specific genre can only emerge over a longer period of time.

This concludes the discussion of the findings and their relevance: I will now move on to the conclusion of this dissertation in Chapter 10.

# Chapter 10

## Conclusion

### 10.1 Summary

This dissertation studied the structure of a common multimodal artefact — the tourist brochure — with the aim of establishing the factors that contribute to the artefact’s multimodal structure. My goal was to produce a model of artefact structure to explain the use of semiotic modes, their interaction and interpretation in the tourist brochures. For this purpose, I collected a data set consisting of 89 brochures produced by the city of Helsinki between the years 1967 and 2008.

A literature review established that a model of artefact structure would require an approach preferring the study of multimodal structure over detailed descriptions of meaning-making. Moreover, the chosen approach would have to be able to produce comparable analyses. I thus chose the *Genre and Multimodality* (GeM) model (Bateman 2008) due to its multiple analytical layers, complemented by an XML annotation schema for creating multimodal corpora. Drawing on the collected data, I annotated and compiled a corpus of 58 double-pages using the GeM model.

The annotated corpus contained information on the content, its layout, appearance and rhetorical structure in the tourist brochures. To explain the data, however, additional theoretical concepts had to be introduced. I therefore applied the notions of medium, semiotic mode and genre to outline the various contributions to the artefact structure. I also brought in additional perspectives to artefact structure from the fields of information design and document theory. For making the artefact structure explicit, I developed methods that combined information from several layers of the GeM-annotated corpus and represented this information using computer-drawn diagrams.

The analysis revealed features of multimodal structure that had not been considered previously. Using the developed methods, I was able to show the functional

motivation behind artefact structure. By focusing on the hierarchical organisation of the content and its presentation to the reader, I uncovered several multimodal structures used for specific communicative tasks, such as describing the destination and guiding the reader. I concluded that these patterned structures (cf. Waller et al. 2012) facilitate the use of the tourist brochures by keeping the content and structure in line with the readers' expectations. In addition, their combinations signal how the structure of the tourist brochure is to be interpreted, thus enabling the readers to access to the actual content.

The in-depth analysis of structure revealed two semiotic modes active in the tourist brochures. The first semiotic mode was text-flow, which used language for extensive descriptions of destination, locations and events, often accompanied by various images. The second semiotic mode was page-flow, which used the space of the entire page to segment the verbal and visual content into fragmented but easily digestible units. The analyses revealed the structural configuration of these semiotic modes, which enabled their longitudinal study. A major finding was that page-flow has become more prominent after the year 1985, which marked the introduction of desktop publishing, that is, computer-assisted graphic design and digital printing.

Finally, the developed model did not only provide a comprehensive description of the tourist brochures and their multimodal structure, but also helped to situate the previously studied multimodal phenomena. Additionally, the model may inform the fields of information and graphic design, which tie in with topical issues such as visual and document literacy. These issues are in need of analytical attention and empirical research, which would allow informed education policy-making in relation to multimodality. However, making this kind of research possible will require larger data sets, whose creation will likely require the input of specialists in computer vision and document recognition.

## 10.2 Evaluation of the study

I will begin by assessing the strengths of the study. The definite strengths of this study were the corpus-based approach and the developed methods, which revealed many aspects of artefact structure that the previous analyses had not considered (see e.g. Matthiessen 2007; Yang 2008; Velasco 2012; Taboada and Habel 2013; Kong 2013). The structured and cross-referenced data stored in the GeM corpus enabled the so-called cross-layer analyses, which combined data from multiple analytical layers. Using these cross-layer analyses, I identified several configurations of semiotic modes that reside within the intertwined layout and rhetorical structures of multimodal artefacts (see Hiippala 2013).

Moreover, the corpus-based approach allowed the analytical focus to be shifted according to need, in the way envisaged by Flewitt et al. (2009, p. 44):

The combination of ‘new’ and more traditional tools for data collection creates a dynamic constellation of resources, where meanings are produced through the inter-relationships between and within the data sets, permitting the researcher literally and metaphorically to ‘zoom in’ on fine-grained detail and to pan out to gain a broader, socially and culturally situated perspective.

The combination of traditional query-based methods of corpus linguistics with novel visualisation techniques offered a more precise view of the multimodal structure, especially in terms of tracking semiotic change (cf. Bezemer and Kress 2009). At the same time, the study has certain potential weaknesses, which I will outline next.

Throughout this study, I have emphasised the importance of an empirical approach to multimodal analysis. For this reason, the potential weaknesses of this study need to be evaluated in this context. In practice, this means that the weaknesses are mainly concerned with empirical validity, and these issues are naturally related to the creation of the annotated corpus (see Section 5.1.3.2).

The selection of the annotated corpus could have been based on a probabilistic sample (cf. Biber 1993). Although the annotated corpus consisted of a random sample of pages from the entire data set, I consciously included examples from each brochure series in the data. With a more careful pre-processing of the data, the selection could have been turned into a probabilistic sample by identifying the page types in the data and sampling these types randomly. Paradoxically, such an understanding of the page types in the tourist brochures was not available until the completion of this study.

The issue of annotating the GeM corpus should also be considered. I observed the development of my knowledge of the GeM annotation schema during the three and half years of working with the corpus. This became evident during the verification of the corpus, as more errors were found in the early annotation work (see Section 5.4). Moreover, the time spent annotating the corpus also shows why optical character recognition technology needs to be brought to bear on the creation of multimodal corpora in the near future (Thomas 2009*b*, p. 230). Although the creation of the annotated corpus was a time-consuming process, this investment was undoubtedly worthwhile in the light of the results.

The final issue related to the corpus and the empirical validity of this study is the fact that the work was carried out by one annotator only. Annotator agreement has been previously measured and studied in RST research (see Taboada and Mann 2006*b*, pp. 443-445) and these techniques are still being developed (cf. Mitocariu et al. 2013). These techniques could have been adopted in this dissertation as well,

but the time-consuming nature of the annotation work and the lack of specialists in this area prevented their implementation.

## 10.3 Implications

In this section, I outline the implications of this dissertation. I shall first discuss the methodological and theoretical implications of the research in Section 10.3.1. I then present how the major findings of this dissertation may be applied to tourist brochure design in Section 10.3.2.

### 10.3.1 Contributions to multimodal research

According to Bateman (2008, p. 8), the GeM model aimed to “provide a foundation for an investigative method that is sufficiently robust to advance theory empirically”. Forceville (2010, p. 2607), in turn, underlined the importance of carefully circumscribed multimodal corpora for sharpening and refining the analytic tools used in multimodal research. By combining these two perspectives, this dissertation showed that a corpus-based approach is indeed a step towards a more empirically-responsible direction in multimodal research. To support my argument, I will now make the contributions of this dissertation explicit below.

Methodologically, the dissertation showed the benefits of data-driven analysis in multimodal research. With the annotated corpus at hand, the shift of perspective upon encountering analytical challenges became possible. Such shifts were shown for the layout layer in Section 7.1.1 and for the navigation layer in Section 7.3.3. This illustrated the flexibility of a corpus-based approach: when a particular layer of the GeM model was not found relevant for the aspect of artefact structure under analysis, an alternative could be immediately sought using the other analytical layers or their combinations.

Visually, the structured data stored in the corpus could be transformed, represented and studied using query and graph description languages, such as XQuery and DOT. These visualisations, which combined information from multiple analytical layers of the GeM model, provided a vantage point to the structure of the artefact and supported its analysis. The abstractions provided by the visualisations allowed the dissertation to look below the surface of the artefact, which could not have been achieved using the traditional corpus linguistic methods such as keyword-in-context lists (cf. Bateman 2012). To sum up, the visualisations helped to identify patterns in the multimodal structure. Without automatisations, this would not have been possible, as the visualisation of the entire annotated corpus required over 10000 lines of code in DOT.

Theoretically, the model of artefact structure advanced our understanding of the structure of a multimodal artefact and helped to situate the previously studied multimodal phenomena. In this way, the model developed in this dissertation provides a framework to support the detailed analyses, as envisaged by Forceville (2007, p. 1236). This represents the first step in forging a stronger bond between the multimodal analyses of content and structure.

In more general terms, a model of artefact structure may eventually provide multimodal research and information design with a yardstick similar to “grammaticality” in linguistics (Waller et al. 2012, p. 4), at least to a certain extent. Although mapping every possible choice related to artefact structure is not feasible at the moment, laying bare the basic principles of multimodal meaning-making may help the field to reach out to the practitioners of graphic and information design. Establishing a mutually beneficial relation between these fields should be based on cooperation, because a theoretically-oriented approach cannot impose its views on a practice-based trade without conflict. Such a development would also strengthen the position of multimodal research among other disciplines focused on communication.

However, establishing multimodal research as a reputable field of study in the eyes of the language and communication scholars is not enough. To establish itself firmly, the field of research needs to show its relevance to the larger audience as well. And the best way to bring out the benefits of multimodal research is to convert the research results into pragmatic advice, which I seek to accomplish in the following section.

### **10.3.2 Pragmatic advice on tourist brochure design**

If multimodal research wants to have an impact on the society at large, we need to be able to translate our findings into a language understood by those without linguistic or multimodal training. For this reason, I will now present the main findings of this dissertation in plain language. Specifically, I attempt to outline “the most relevant attributes in brochure design in general and their main attractiveness, content, and usefulness features in particular”, which Molina and Esteban (2006, p. 1051) identified as an open issue in their study of tourist brochures. Therefore, I will now address two issues, which may be roughly divided along the lines of image formation and brochure usefulness.

According to Molina and Esteban (2006, p. 1047), the “visual format” of the brochures is strongly connected to the process of building an image of the destination. For this reason, I shall first focus on the forms of visual communication in the tourist brochures.

In my analysis, I established that the tourist brochures use two strategies for incorporating visual content into the brochures. The first strategy is to provide the

images with captions, thus establishing a direct link between the verbal and visual content. In this way, the images are not left open for interpretation in relation to the other content of the brochure: this strategy is often used to describe specific locations or events within a destination.

The second strategy involves what I termed ‘conceptual’ images. These photographs do not establish a direct link to the text. Although they may be alluded to in the verbal content — which is a common feature of any text combining language and images — no direct connection is established between the text and the accompanying images using captions. Without a direct link to the text, these images are left open for interpretation: they speak for themselves in creating a mental image of the destination.

The two strategies described above can be seen as solutions to different types of communicative needs. The first strategy may be more appropriate for describing a particular location or event to a great detail, providing the reader with a detailed description. The second strategy, in turn, is more likely to be used in building a mental image of the destination. However, satisfaction with the brochures is also dependent on what Molina and Esteban refer to as the “functional attributes” and “informative attractiveness and content” (2006, p. 1048). Therefore, I now turn to discuss what this dissertation revealed about the structure and functions of the tourist brochures.

To begin with, the readers often have expectations towards the structure and content of the tourist brochures, that is, they expect to find certain types of information, communicated in a familiar way. For this reason, the *structure* of the tourist brochures should be designed to meet these expectations. How information is presented to the reader should follow a consistent pattern: these features enable the reader to recognise the tourist brochure, while simultaneously invoking mental models that tell the reader how the brochure should be used.

Detailed descriptions of the destination and guidance are often provided using written language, while the two strategies described above are used to integrate images into the structure of the brochures. Walking routes and guides typically follow a sequenced structure, which is split into subsequences that describe the sights. For representing geographical information, a map is naturally the most efficient way. Together, these points make up the structure of the brochure, which may be thought of as its ‘wireframe’: the visual appearance of the content may be determined by the graphic designer.

For the sake of usability, the structure and appearance of a tourist brochure should follow a clear pattern, as this reduces the need for cognitive processing. In the recent years, the tourist brochure design has moved towards a structure that organises the content visually into digestible units, which allow a large amount of content to be included on each page. Moreover, the traditional linear organisation



of the page is no longer followed and related content may exist on different areas of the page. To enable the reader to cope with increasingly complex designs, the brochures should take care to follow a clearly designated pattern in both structure and content.

## 10.4 Avenues of further research

The research conducted in this dissertation revealed several possible avenues of further research. In this section, I present four possible directions for future research, which all draw on the developed model of artefact structure.

Firstly, the developed model could be used to compare the structure of different multimodal artefacts. Because the tourist brochures are mainly descriptive, an interesting point of comparison could be found, for example, in artefacts whose function is mainly instructive (see e.g. André and Rist 1995; Martinec 2003). A comparative approach could lead to a more informed view of the functionally motivated structure in multimodal artefacts across different cultures (cf. Bateman and Delin 2003; Hiippala 2012*a*). This also presents one possible case of re-use for the GeM-annotated corpus produced in this dissertation.

This leads to the second issue: how to meet a major challenge faced by multimodal corpus linguistics — the corpora. This “logjam” is the creation of multimodal corpora, which optical character recognition helped to solve for the field of linguistics (Leech 1991, p. 10). But as Parodi (2010, p. 72) observes:

... machine-readable digital multimodal automatic text identification lacks a robust theory of (multimodal) language in the framework of the so-called “visual turn.”

This particularly evident in the fields of automatic document recognition and computer vision, in which the definitions of genre remain underdeveloped. At the same time, researchers within these fields point out that “document genres ... can improve document search” and “documents occur in different modalities” (Chen et al. 2012, p. 167). This is undoubtedly an area to which multimodal research can contribute after a common ground is found. In some cases, possible points of contact between the two fields already exist: for instance, the concepts used to describe document structure in Rangoni et al. (2012) are very close to those used for the layout layer of the GeM model.

Thirdly, the model of artefact structure can also inform us about the processes of design and production, which are direly in need of further attention from a multimodal perspective. The researcher is often left only with the end product with little knowledge of its creation. However, the model could be used for mutual benefit in ethnographic investigations (see e.g. Kress 2011) by providing snapshots

of the artefact structure during the unfolding design process. Because this process is computerised, the observation and analysis of artefact structure could be built into the design software. This opens many possibilities for research, which also tie in with the field of graphic design and how it could be taught in the future.

The fourth and final avenue of further research suggested here is arguably also the most critical one. This is the issue of “visual turn” in communication, which Parodi (2010) also mentions in the quote above. In this dissertation, I showed that the visual turn may be more complex than previous research has suggested. Consequently, the concept may need to be reworked, as the visual turn does not necessarily manifest itself only in terms of visibility, but also in terms of the artefact’s multimodal structure and the configuration of the semiotic modes.

In particular, an understanding of the visual turn and contemporary communication are highly important to future education. While teaching the critical interpretation of content is certainly useful and necessary, it is equally important to consider how an understanding of multimodal structure may help the learners to access this content. This understanding, however, can only be built on a robust empirical foundation — which is also a prerequisite for informed policy-making in education. To achieve this, a close coupling needs to be established between the theories of multimodality and the analysed data.

# Appendix A

## The entire data set

This appendix provides the entire data set collected for the dissertation. The data are given in Table A.1. The column headings are explained below:

1. **Brochure name:** The name of the brochure.
2. **Year:** The brochure's year of publication.
3. **S:** The source of the data, using the following abbreviations:
  - *H*: The Helsinki City Archives.
  - *N*: The National Library of Finland
  - *O*: The author's personal collection.
4. **D:** The number of double-pages in the brochure.
5. **P:** Whether the brochure includes page numbering, (Y)es or (N)o.
6. **I:** Whether the brochure includes an index, (Y)es or (N)o.
7. **F:** Whether the pages of the brochure are folded, (Y)es or (N)o.
8. **B:** Whether the brochure is bound using staples, (Y)es or (N)o.  
It should be noted that some brochures are both folded *and* bound.
9. **Cp:** The number of content pages in the brochure. The count excludes empty pages and pages with purely navigational content, such as indices.
10. **Fs:** The number of folds per page. Mainly relevant for the folded brochures.
11. **A:** Whether the brochure includes classified advertisements, (Y)es or (N)o.

Table A.1: The entire data set

Brochure name	Year	S	Dp	Pn	I	F	B	Cp	Fs	Ads
Helsinki's Four Tourist Islands	1967	H	2	N	N	Y	N	4	3	N
Museums and exhibitions in Helsinki (Issue 7)	1967	H	2	N	N	Y	N	4	2	N
Sculptures and Monuments in Helsinki	1967	H	2	Y	N	Y	N	4	2	N
Coming Events in Helsinki (Issue 21)	1967	H	2	Y	N	Y	N	4	2	N
Coming Events in Helsinki (Issue 22)	1967	H	2	Y	N	Y	N	4	2	N
Helsinki Tourist Information	1968	H	2	Y	N	Y	N	8	4	N
Museums and Exhibitions in Helsinki	1968	H	9	N	N	N	Y	10	1	N
Museums and Exhibitions in Helsinki (Issue 9)	1968	H	2	Y	N	Y	N	4	2	N
Museums and Exhibitions in Helsinki (Issue 10)	1968	H	2	Y	N	Y	N	4	2	N
Sculptures and Monuments in Helsinki	1968	H	2	Y	N	Y	N	4	2	N
Helsinki Daughter of the Baltic	1969	H	2	N	N	Y	N	4	3	N
Helsinki Tourist Information	1971	H	2	Y	N	Y	N	10	4	N
Helsinki Tourist Information	1971	H	2	Y	Y	Y	N	8	4	N
Coming Events in Helsinki (Issue 46)	1971	H	2	N	N	Y	N	4	2	N
Coming Events in Helsinki (Issue 48)	1971	H	2	N	N	Y	N	4	2	N
Helsinki Daughter of the Baltic	1972	H	2	N	N	Y	N	14	5	N
Helsinki Tourist Information	1972	H	10	Y	Y	N	Y	17	1	N
Helsinki's Four Tourist Islands	1972	H	2	N	N	Y	N	4	3	N
Sights in and around Helsinki	1972	H	10	Y	Y	N	Y	17	1	N
Helsinki Daughter of the Baltic	1973	H	2	N	N	Y	N	14	5	N
Helsinki Tourist Information	1974	H	6	Y	Y	N	Y	9	1	N
Sculptures and Monuments in Helsinki	1974	H	10	Y	Y	N	Y	18	1	N
Helsinki Daughter of the Baltic	1975	H	2	N	N	Y	N	14	5	N
Helsinki's Four Tourist Islands	1976	H	6	N	N	N	Y	11	3	N
Helsinki's Four Tourist Islands (Issue 2)	1976	H	6	N	N	N	Y	11	3	N
Sights in and around Helsinki	1976	H	11	Y	Y	N	Y	20	1	N
Helsinki Daughter of the Baltic	1977	H	2	N	N	Y	N	14	5	N
Helsinki Tourist Information	1977	H	6	Y	Y	N	Y	9	1	N
A Walking Tour in Helsinki (Eläintarha-Töölö)	1978	H	2	N	N	Y	N	2	1	N
A Walking Tour in Helsinki (Katajanokka-Esplanadi-Tähtitorninmäki)	1978	H	2	N	N	Y	N	3	1	N
A Walking Tour in Helsinki (Kruununhaka)	1978	H	2	N	N	Y	N	2	1	N
Helsinki's Four Tourist Islands	1979	H	6	N	N	N	Y	11	3	N
Helsinki Daughter of the Baltic	1980	H	2	N	N	Y	N	14	5	N
The Greater Helsinki	1980	H	8	N	N	N	Y	14	3	N
Helsinki Daughter of the Baltic	1981	H	2	N	N	Y	N	14	5	N
Sculptures and Monuments in Helsinki	1982	H	7	Y	Y	N	Y	11	1	N
A Walking Tour in Helsinki (Eläintarha-Töölö)	1983	H	2	N	N	Y	N	2	1	N
A Walking Tour in Helsinki (Kaivopuisto-Eira)	1983	H	2	N	N	Y	N	3	1	N

A Walking Tour in Helsinki (Katajanokka-Esplanade)	1983	H	2	N	N	Y	N	3	1	N
A Walking Tour in Helsinki (Kruununhaka)	1983	H	2	N	N	Y	N	3	1	N
Helsinki Daughter of the Baltic	1983	H	2	N	N	Y	N	14	5	N
Holiday in Helsinki	1983	H	12	Y	Y	N	Y	20	1	N
Helsinki Daughter of the Baltic	1984	H	2	N	N	Y	N	14	5	N
Helsinki's Four Tourist Islands	1984	H	6	N	N	N	Y	11	3	N
Holiday in Helsinki	1984	H	12	Y	Y	N	Y	20	1	N
Suomenlinna Seafortress in Helsinki	1984	H	2	N	N	Y	N	8	3	N
Helsinki Daughter of the Baltic	1985	H	2	N	N	Y	N	14	5	N
Holiday in Helsinki	1985	H	12	Y	Y	N	Y	20	1	N
Suomenlinna Seafortress in Helsinki	1985	H	2	N	N	Y	N	8	3	N
Helsinki Daughter of the Baltic	1986	H	2	N	N	Y	N	14	5	N
Helsinki's Four Tourist Islands	1986	H	6	N	N	N	Y	11	3	N
Holiday in Helsinki	1986	H	12	Y	Y	N	Y	19	1	N
Suomenlinna Seafortress in Helsinki	1986	H	2	N	N	Y	N	8	3	N
Holiday in Helsinki	1987	H	12	Y	Y	N	Y	19	1	N
See Helsinki on Foot	1987	H	14	N	N	N	Y	25	3	N
Suomenlinna Seafortress in Helsinki	1987	H	2	N	N	Y	N	8	3	N
Helsinki Daughter of the Baltic	1988	H	2	N	N	Y	N	14	5	N
Helsinki's Four Tourist Islands	1988	H	6	N	N	N	Y	11	3	N
Holiday in Helsinki	1988	N	14	Y	Y	N	Y	23	1	N
See Helsinki on Foot	1988	N	14	N	Y	N	Y	25	3	N
Suomenlinna Seafortress in Helsinki	1988	H	2	N	N	Y	N	8	3	N
Holiday in Helsinki	1989	N	14	Y	Y	N	Y	23	1	N
See Helsinki on Foot	1990	N	14	N	Y	N	Y	25	3	N
See Helsinki on Foot	1991	N	14	N	Y	N	Y	25	3	N
Holiday in Helsinki	1992	N	8	Y	Y	N	Y	12	3	N
See Helsinki on Foot	1994	N	14	N	Y	N	Y	25	3	N
Welcome to Helsinki	1995	N	10	Y	Y	N	Y	17	1	N
Welcome to Helsinki	1996	N	10	Y	Y	N	Y	17	1	N
Welcome to Helsinki	1997	N	10	Y	Y	N	Y	17	1	N
Welcome to Helsinki	1998	N	10	Y	Y	N	Y	17	1	N
Helsinki Your Way	1999	N	32	Y	Y	N	Y	61	1	N
Helsinki What Where When	1999	N	32	Y	Y	N	Y	60	1	N
Helsinki 1550-2000	2000	N	2	N	N	Y	N	4	5	N
Suomenlinna Maritime Fortress	2001	H	10	Y	Y	N	Y	17	1	N
Summer-Helsinki	2002	H	26	Y	Y	N	Y	49	1	Y
Helsinki Your Way	2003	H	32	Y	Y	N	Y	61	1	Y
Winter Helsinki	2003	H	22	Y	Y	N	Y	35	1	N
Helsinki Your Way	2005	H	32	Y	Y	N	Y	59	1	Y
See Helsinki on Foot	2005	H	13	N	N	N	Y	25	1	N
Bohemian Nordic Oddity	2006	O	2	N	N	Y	N	3	2	N
Groovy Nordic Oddity	2006	O	2	N	N	Y	N	3	2	N
Helsinki Your Way	2006	O	32	Y	Y	N	Y	60	1	Y
Smooth Nordic Oddity	2006	O	2	N	N	Y	N	3	2	N
Weekend Delight in Helsinki	2006	O	8	N	N	N	Y	14	1	Y
See Helsinki on Foot	2007	O	20	Y	Y	N	Y	38	1	N
Bohemian Nordic Oddity	2008	O	2	N	N	Y	N	3	2	N

Groovy Nordic Oddity	2008	O	2	N	N	Y	N	3	2	N
Helsinki Visitors Guide	2008	O	33	Y	Y	N	Y	67	1	Y
Smooth Nordic Oddity	2008	O	2	N	N	Y	N	3	2	N

---

# Appendix B

## The rhetorical relations used in GeM RST

Tables B.1, B.2 and B.3 list the rhetorical relations defined in the GeM annotation schema: the abbreviated descriptions in the tables are based on Mann (2005) and Bateman (2008). The following abbreviations are used in the tables: *N* (nucleus), *S* (satellite), *R* (reader), *W* (writer). The first column gives the relation name; the second defines the constraints on either nucleus or satellite. The third column gives constraints that affect the combination of nucleus and satellite. The fourth column describes the intended outcome of the rhetorical relation from the writer’s perspective.

Table B.1: List of asymmetric RST relations in the GeM RNG DTD

Relation	Constraints on N/S	Constraints on N+S	Outcome
ANTITHESIS	W has positive regard for N.	W cannot have positive regard for both N & S.	R’s positive regard for N is increased.
BACKGROUND	R won’t comprehend N before reading S.	S increases the ability of R to comprehend an element in N.	R’s ability to comprehend N increases.
CIRCUMSTANCE	S is not unrealised.	S provides the framework within which R is to interpret N.	R recognises that S provides the framework for interpreting N.
CONCESSION	N: W has positive regard for N. S: W is not claiming that S does not hold.	W acknowledges a potential or apparent incompatibility between N and S.	R’s positive regard for N is increased.
CONDITION	S presents a hypothetical, future, or otherwise unrealised situation.	Realisation of N depends on realisation of S.	R recognises how the realisation of N depends on S.

ELABORATION	None.	S presents additional detail about N.	R recognises that S provides additional information on N.
ENABLEMENT	N presents an unrealised action by R.	S enables R to perform the action described in N.	R's capability to perform the action in N increases.
EVALUATION	None.	S relates N to degree of W's positive regard toward N.	R recognises that S assesses N and recognises the value it assigns.
EVIDENCE	N: R might not believe N. S: R finds S credible.	R's comprehending of S increases R's belief of N.	R's belief of N is increased.
INTERPRETATION	None.	S relates N to a framework of ideas not involved in N itself and not concerned with W's positive regard.	R recognizes that S relates N to a framework of ideas not involved in the knowledge presented in N itself.
JUSTIFY	None.	S increases R's readiness to accept W's right to present N.	R's readiness to accept W's right to present N is increased.
MEANS	N is an activity.	S makes the realisation of N more likely.	R recognises that S makes the realisation of N more likely.
MOTIVATION	N is an action in which R is the actor.	Comprehending S increases R's desire to perform action in N.	R's desire to perform action in N is increased.
NONVOLITIONAL-CAUSE	N is not a volitional action.	S, by means other than motivating a volitional action, caused N.	R recognises S as a cause of N.
NONVOLITIONAL-RESULT	S is not a volitional action.	N caused S.	R recognizes that N could have caused the situation in S.
OTHERWISE	N is an unrealised situation. S is an unrealised situation.	Realisation of N prevents realisation of S.	R recognises the relation of realisation between N & S.
PREPARATION	None.	S precedes N in the text; S prepares R for N.	R is more ready, interested or oriented for reading N.
PURPOSE	N is an activity; S is an unrealised situation.	S is realised through the activity in N.	R recognises that the activity in N is initiated in order to realise S.
RESTATEMENT	None.	S restates N.	R recognises S as a restatement of N.



SOLUTIONHOOD	S presents a problem.	N is a solution to the problem presented in S.	R recognises N as a solution to S.
SUMMARY	N must be more than one unit.	S presents a shorter re-statement of N.	R recognises S as a shorter restatement of N.
UNLESS	None.	S affects the realisation of N; N is realised if S is not.	R recognises that N is realised if S is not.
VOLITIONAL-CAUSE	N is a volitional action.	S could have caused the agent of the volitional action in N to perform that action.	R recognises S as a cause for N.
VOLITIONAL-RESULT	S is a volitional action.	N could have caused S.	R recognises that N could be a cause for S.

Table B.2: List of symmetric RST relations in the GeM RNG DTD

Relation	Constraints on N+N	Outcome
CONTRAST	No more than two nuclei.	Reader acknowledges the comparability of the nuclei.
JOINT	None.	None. There are no constraints on assigning the relation between nuclei.
LIST	An item comparable to others linked to it by the LIST relation.	R recognises the comparability of linked items.
RESTATEMENT	The nuclei re-express each other.	R recognises the re-expression.
SEQUENCE	There is a succession relationship between the nuclei.	R recognises the succession.

Table B.3: List of subnuclear RST relations in the GeM RNG DTD

Relation	Description
IDENTIFICATION	Assertion of identity.
CLASS-ASCRPTION	Relation between an object and its superclass.
PROPERTY-ASCRPTION	Relation between an object and something predicated of that object.
POSSESSION	Relation between possessor and possessed.
LOCATION	Relation between an object and its spatial or temporal location.
PROJECTION	Undocumented!

(Source: Bateman 2008, p. 162)

# Bibliography

- Aiello, G. (2006), 'Theoretical advances in critical visual analysis: Perception, ideology, mythologies, and social semiotics', *Journal of Visual Literacy* **26**(2), 89–102.
- Alias, S. (2004), A semiotic study of Singapore's Orchard Road and Marriott Hotel, in K. L. O'Halloran, ed., 'Multimodal Discourse Analysis: Systemic Functional Perspectives', Continuum, London, pp. 55–79.
- Allen, P. T., Bateman, J. A. and Delin, J. L. (1999), Genre and layout in multimodal documents: towards an empirical account, in R. Power and S. Donia, eds, 'Using Layout for the Generation, Understanding or Retrieval of Documents. Proceedings of the American Association for Artificial Intelligence Fall Symposium', AAAI Press, Menlo Park, CA, pp. 27–34.
- Allwood, J. (2008), Multimodal corpora, in A. Lüdeling and M. Kytö, eds, 'Corpus Linguistics: An International Handbook', Mouton de Gruyter, Berlin, pp. 207–225.
- Altman, R. (1999), *Film/Genre*, British Film Institute, London.
- Andersen, J. (2009), 'The concept of genre in information studies', *Annual Review of Information Science and Technology* **42**(1), 339–367.
- André, E. and Rist, T. (1995), 'Generating coherent presentations employing textual and visual material', *Artificial Intelligence Review* **9**, 147–165.
- Asher, N. and Lascarides, A. (2003), *Logics of Conversation*, Cambridge University Press, Cambridge.
- Astorga, M. C. (2009), The text-image matching: One story, two textualizations, in E. Ventola and A. J. Guijarro, eds, 'The World Told and the World Shown: Multi-semiotic Issues', Palgrave, London, pp. 124–136.
- Baldry, A. P. (2000a), ESP in a visual society: historical dimensions in multimodality and multimediality, in A. P. Baldry, ed., 'Multimodality and Multimediality in the Distance Learning Age', Palladino, Campobasso, pp. 41–89.
- Baldry, A. P. (2007), The role of multimodal concordancers in multimodal corpus linguistics, in T. D. Royce and W. L. Bowcher, eds, 'New Directions in the Analysis of Multimodal Discourse', Lawrence Erlbaum, Mahwah, NJ, pp. 173–193.
- Baldry, A. P., ed. (2000b), *Multimodality and Multimediality in the Distance Learning Age*, Palladino, Campobasso.
- Baldry, A. P. and Montagna, E., eds (2011), *Interdisciplinary approaches to multimodality: Theory and practice*, Palladino, Campobasso.

- Baldry, A. P. and Thibault, P. J. (2005), *Multimodal transcription and text analysis*, Equinox, London.
- Barthes, R. (1977), *Image-Music-Text: Essays*, Hill and Wang, New York. Translated by Stephen Heath.
- Barthes, R. (1981), *Camera Lucida*, Hill and Wang, New York. Translated by Richard Howard.
- Bateman, J. A. (2007), 'Towards a *grande paradigmatique* of film: Christian Metz reloaded', *Semiotica* **167**(1), 13–64.
- Bateman, J. A. (2008), *Multimodality and Genre: A Foundation for the Systematic Analysis of Multimodal Documents*, Palgrave Macmillan, London.
- Bateman, J. A. (2009a), 'The decomposability of semiotic modes', MS University of Bremen.
- Bateman, J. A. (2009b), Discourse across semiotic modes, in J. Renkema, ed., 'Discourse, of course: An overview of research in discourse studies', Benjamins, Amsterdam/Philadelphia, pp. 55–66.
- Bateman, J. A. (2009c), Film and representation: making filmic meaning, in W. Wildgen and B. van Heusden, eds, 'Metarepresentation, Self-Organization and Art', European Semiotics, Peter Lang, Bern, pp. 137–162.
- Bateman, J. A. (2011), The decomposability of semiotic modes, in K. L. O'Halloran and B. A. Smith, eds, 'Multimodal Studies: Multiple Approaches and Domains', Routledge, London, pp. 17–38.
- Bateman, J. A. (2012), Multimodal corpus-based approaches, in C. A. Chapelle, ed., 'The Encyclopedia of Applied Linguistics', Wiley-Blackwell, Oxford, pp. 3983–3991.
- Bateman, J. A. (forthcoming), Genre in the age of multimodality: some conceptual refinements for practical analysis, in P. Evangelisti-Allori, V. K. Bhatia and J. A. Bateman, eds, 'Evolution in Genres: Emergence, Variation, Multimodality', Peter Lang, Frankfurt am Main.
- Bateman, J. A. and Delin, J. L. (2003), Genre and multimodality: expanding the context for comparison across languages, in D. Willems, B. Defrancq, T. Coleman and D. Noël, eds, 'Contrastive Analysis in Language: Identifying Linguistic Units of Comparison', Palgrave Macmillan, Houndsmill, pp. 230–266.
- Bateman, J. A., Delin, J. L. and Allen, P. (2000), Constraints on layout in multimodal document generation, in 'Proceedings of the First International Natural Language Generation Conference, Workshop on Coherence in Generated Multimedia', Association for Computational Linguistics, Mitzpe Ramon, Israel.  
**URL:** <http://www.hcrc.ed.ac.uk/~john/inlg-mm/papers/bateman-final.pdf>
- Bateman, J. A., Delin, J. L. and Henschel, R. (2002a), 'Multimodality and empiricism: methodological issues in the study of multimodal meaning-making. gem project report', MS University of Bremen and University of Stirling.
- Bateman, J. A., Delin, J. L. and Henschel, R. (2002b), XML and multimodal corpus design: experiences with multi-layered stand-off annotations in the GeM corpus, in 'Proceedings of the LREC'02 Workshop 'Towards a Roadmap for Multimodal Language Resources and Evaluation'', Las Palmas, Canary Islands, Spain.

**URL:** <http://www.elsnet.org/dox/lrec2002-bateman.pdf>

- Bateman, J. A., Delin, J. L. and Henschel, R. (2004), Multimodality and empiricism: preparing for a corpus-based approach to the study of multimodal meaning-making, in E. Ventola, C. Charles and M. Kaltenbacher, eds, 'Perspectives on Multimodality', Benjamins, Amsterdam, pp. 65–89.
- Bateman, J. A., Delin, J. L. and Henschel, R. (2007), Mapping the multimodal genres of traditional and electronic newspapers, in T. D. Royce and W. L. Bowcher, eds, 'New Directions in the Analysis of Multimodal Discourse', Lawrence Erlbaum, Mahwah, NJ, pp. 147–172.
- Bateman, J. A. and Henschel, R. (2007), Generating text, layout and diagrams appropriately for genre, in I. van der Sluis, M. Theune, E. Reiter and E. Krahmer, eds, 'Proceedings of the Workshop on Multimodal Output Generation MOG 2007', Centre for Telematics and Information Technology (CTIT), University of Twente, pp. 29–40.
- Bateman, J. A. and Rondhuis, K. (1997), 'Coherence relations: Towards a general specification', *Discourse Processes* **24**(1), 3–49.
- Bateman, J. A. and Schmidt, K.-H. (2012), *Multimodal Film Analysis: How Films Mean*, Routledge, London.
- Bawarshi, A. S. and Reiff, M. J. (2010), *Genre: An Introduction to History, Theory, Research, and Pedagogy*, Parlor, West Lafayette.
- Bazerman, C. (1988), *Shaping Written Knowledge: The Genre and Activity of the Experimental Article in Science*, The University of Wisconsin Press, Madison.
- Beasley, R. and Danesi, M. (2002), *Persuasive Signs: The Semiotics of Advertising*, Mouton de Gruyter, Berlin & New York.
- Bednarek, M. and Martin, J. R. (2010), *New Discourse on Language: Functional Perspectives on Multimodality, Identity, and Affiliation*, Continuum, London.
- Bell, P. (2001), Content analysis of visual images, in T. van Leeuwen and C. Jewitt, eds, 'Handbook of Visual Analysis', SAGE Publications, London, pp. 10–34.
- Bell, P. and Milic, M. (2002), 'Goffman's gender advertisements revisited: combining content analysis with semiotic analysis', *Visual Communication* **1**(2), 203–222.
- Berazhny, I. (2008), Construing an airline destination: English clause in a network, in C. Wu, C. M. Matthiessen and M. Herke, eds, 'Proceedings of ISFC 35: Voices Around the World', Macquarie University, The 35th ISFC organizing committee, Sydney, pp. 284–290.
- Berazhny, I. (2012), Alcohol advertising in *Aeroflot*, in W. L. Bowcher, ed., 'Multimodal Texts from Around the World: Linguistic and Cultural Insights', Palgrave, London, pp. 146–172.
- Berazhny, I. and Hiippala, T. (in preparation), 'Communicating corporate social responsibility in in-flight magazines'.
- Berger, A. A. (2011), 'Tourism as a postmodern semiotic activity', *Semiotica* **183**(1), 105–119.
- Berkenkotter, C. and Huckin, T. N. (1995), *Genre Knowledge in Disciplinary Communication: Cognition, Culture, Power*, Lawrence Erlbaum, Hillsdale, N.J.

- Bezemer, J. and Kress, G. (2009), 'Visualizing English: a social semiotic history of a school subject', *Visual Communication* **8**(3), 247–262.
- Bhatia, V. K. (1993), *Analysing Genre: Language Use in Professional Settings*, Longman, London.
- Biber, D. (1988), *Variation across speech and writing*, Cambridge University Press, Cambridge.
- Biber, D. (1993), 'Representativeness in corpus design', *Literary & Linguistic Computing* **8**(4), 243–257.
- Boeriis, M. and Holsanova, J. (2012), 'Tracking visual segmentation: connecting semiotic and cognitive perspectives', *Visual Communication* **11**(3), 259–281.
- Bonelli, E. T. and Manca, E. (2002), 'Welcoming children, pets and guests: A problem of non-equivalence in the languages of "agriturismo" and "farmhouse holidays"', *Textus* **15**(2), 307–333.
- Bowcher, W. L. (2012), Multimodality in Japanese anti-war placards, in W. L. Bowcher, ed., 'Multimodal Texts from Around the World: Cultural and Linguistic Insights', Palgrave, London, pp. 217–245.
- Bowcher, W. L. and Liang, J. Y. (2013), 'Chinese tourist site entry tickets: intersemiotic complementarity in an ecosocial process', *Social Semiotics* .
- Bunton, D. (2005), 'The structure of PhD conclusion chapters', *Journal of English for Academic Purposes* **4**(3), 207–224.
- Butler, C. (1985), *Systemic Linguistics: Theory and Applications*, Batsford, London.
- Butt, D., Spinks, S. and Yallop, C. (2000), *Using Functional Grammar: An Explorer's Guide*, second edn, National Centre for English Language Teaching and Research, Macquarie University, Sydney.
- Caple, H. (2009a), Multi-semiotic communication in an Australian broadsheet: A new news story genre, in C. Bazerman, A. Bonini and D. Figueiredo, eds, 'Genre in a Changing World', Parlor, West Lafayette, pp. 243–254.
- Caple, H. (2009b), Playing with words and pictures: Intersemiosis in a new genre of news reportage, PhD thesis, University of Sydney.
- Caple, H. and Bednarek, M. (2010), 'Double-take: Unpacking the play in the image-nuclear news story', *Visual Communication* **9**(2), 211–229.
- Catenaccio, P. (2008), 'Press releases as a hybrid genre: Addressing the informative/promotional conundrum', *Pragmatics* **18**(1), 9–31.
- Chambers, J. K., Trudgill, P. and Schilling-Estes, N., eds (2002), *The handbook of language variation and change*, Blackwell handbooks in linguistics, Blackwell, Oxford.
- Chen, F., Girgensohn, A., Cooper, M., Lu, Y. and Filby, G. (2012), 'Genre identification for office document search and browsing', *International Journal on Document Analysis and Recognition* **15**(3), 167–182.
- Chen, Y. (2010a), 'Exploring dialogic engagement with readers in multimodal EFL textbooks in China', *Visual Communication* **9**(4), 485–506.
- Chen, Y. (2010b), 'The semiotic construal of attitudinal curriculum goals: Evidence from EFL textbooks in China', *Linguistics and Education* **21**, 60–74.

- Cheong, Y. Y. (2004), The construal of ideational meaning in print advertisements, *in* K. L. O'Halloran, ed., 'Multimodal Discourse Analysis: Systemic Functional Perspectives', Continuum, London, pp. 163–195.
- Chomsky, N. (1957), *Syntactic Structures*, Mouton, The Hague.
- Chomsky, N. (1995), *The Minimalist Program*, MIT Press, Cambridge, MA.
- Chomsky, N. (2006), *Language and Mind*, third edn, Cambridge University Press, Cambridge.
- Christie, F. and Martin, J. R., eds (1997), *Genre and Institutions: Social Processes in the Workplace and School*, Continuum, London.
- Cohen, D. J. and Snowden, J. L. (2008), 'The relations between document familiarity, frequency, and prevalence and document literacy performance among adult readers', *Reading Research Quarterly* **43**(1), 9–26.
- Constantinou, O. (2005), 'Multimodal discourse analysis: Media, modes and technologies', *Journal of Sociolinguistics* **9**(4), 602–618.
- Crook, J. (2004), 'On covert communication in advertising', *Journal of Pragmatics* **36**(4), 715–738.
- Crystal, D. (2001), *Language and the Internet*, Cambridge University Press, Cambridge.
- Culler, J. (1988), *Framing the Sign: Criticism and its Institutions*, Blackwell, Oxford.
- Delin, J. L. and Bateman, J. A. (2002), 'Describing and critiquing multimodal documents', *Document Design* **3**(2), 140–155.
- Dicks, B., Flewitt, R., Lancaster, L. and Pahl, K. (2011), 'Multimodality and ethnography: working at the intersection', *Qualitative Research* **11**(3), 227–237.
- Djonov, E. N. (2007), 'Website hierarchy and the interaction between content organization, webpage and navigation design: A systemic functional hypermedia discourse analysis perspective', *Information Design Journal* **15**(2), 144–162.
- Djonov, E. N. (2008), Children's website structure and navigation, *in* L. Unsworth, ed., 'Multimodal Semiotics: Functional Analysis in Contexts of Education', Continuum, London, pp. 216–236.
- Djonov, E. N. and van Leeuwen, T. (2011a), 'The semiotics of texture: from tactile to visual', *Visual Communication* **10**(4), 541–564.
- Djonov, E. and van Leeuwen, T. (2011b), Normativity and software: A multimodal social semiotic approach, *in* S. Norris, ed., 'Multimodality in Practice: Investigating Theory-in-practice-through-methodology', Routledge, London, pp. 119–137.
- Dreyfus, S., Hood, S. and Stenglin, M., eds (2010), *Semiotic Margins: Meaning in Multimodalities*, Continuum, London.
- Duncum, P. (2004), 'Visual culture isn't just visual: Multiliteracy, multimodality and meaning', *Studies in Art Education* **45**(3), 252–264.
- E, Marissa K. L., O'Halloran, K. L. and Judd, K. (2011), 'Working at cross-purposes: multiple producers and text-image relations', *Text & Talk* **31**(5), 579–600.
- Edelheim, J. R. (2007), 'Hidden messages: A polysemic reading of tourism brochures', *Journal of Vacation Marketing* **13**(1), 5–17.

- Eggins, S. (2004), *An Introduction to Systemic Functional Linguistics*, second edn, Continuum, London.
- Eggins, S. and Iedema, R. (1997), Difference without diversity: semantic orientation and ideology in competing women's magazines, in R. Wodak, ed., 'Gender Studies', SAGE Publications, London, pp. 165–196.
- Eiler, M. A. (1987), Semiotics of document design, in R. Steele and T. Threadgold, eds, 'Language Topics: Essays in honour of Michael Halliday', Vol. 2, Benjamins, Amsterdam, pp. 461–479.
- Eisenlauer, V. J. (2011), Multimodality and social actions in 'personal publishing' text: From the German 'poetry album' to Web 2.0 'social network sites', in K. L. O'Halloran and B. A. Smith, eds, 'Multimodal Studies: Multiple Approaches and Domains', Routledge, London, pp. 131–152.
- Elleström, L. (2010a), The modalities of media: A model for understanding intermedial relations, in L. Elleström, ed., 'Media Borders, Multimodality and Intermediality', Palgrave, London, pp. 11–48.
- Elleström, L., ed. (2010b), *Media Borders, Multimodality and Intermediality*, Palgrave, London.
- Enkvist, N. E. (1991), Discourse strategies and discourse types, in E. Ventola, ed., 'Functional and Systemic Linguistics: Approaches and Uses', Mouton de Gruyter, Berlin & New York, pp. 3–22.
- Fairclough, N. (1995), *Media Discourse*, Arnold, London.
- Fairclough, N. (2003), *Analysing Discourse: Textual Analysis for Social Research*, Routledge, London.
- Fawcett, R. P. (2000), *A Theory of Syntax for Systemic Functional Linguistics*, Vol. 206 of *Current Issues in Linguistic Theory*, John Benjamins, Amsterdam.
- Feng, D. (2011), Visual space and ideology: A critical cognitive analysis of spatial orientations in advertising, in K. L. O'Halloran and B. A. Smith, eds, 'Multimodal Studies: Multiple Approaches and Domains', Routledge, London, pp. 55–75.
- Firth, J. R. (1957), *Papers in Linguistics, 1931-1951*, Oxford University Press, London.
- Flewitt, R., Hampel, R., Hauck, M. and Lancaster, L. (2009), What are multimodal data and transcription?, in C. Jewitt, ed., 'The Routledge Handbook of Multimodal Analysis', Routledge, London, pp. 40–53.
- Forceville, C. (1996), *Pictorial Metaphor in Advertising*, Routledge, London.
- Forceville, C. (1999), 'Educating the eye? Kress and van Leeuwen's Reading Images: The Grammar of Visual Design', *Language and Literature* 8(2), 163–178.
- Forceville, C. (2007), 'Book review: Multimodal Transcription and Text Analysis: A Multimedia Toolkit and Coursebook', *Journal of Pragmatics* 39(6), 1235–1238.
- Forceville, C. (2010), 'Book review: The Routledge Handbook of Multimodal Analysis', *Journal of Pragmatics* 42(9), 2604–2608.
- Forceville, C. (2011a), 'Book review: Multimodality: A Social Semiotic Approach to Contemporary Communication', *Journal of Pragmatics* 43(14), 3624–3626.

- Forceville, C. (2011b), 'Pictorial runes in *Tintin and the Picaros*', *Journal of Pragmatics* **43**(3), 875–890.
- Fowler, A. (1982), *Kinds of Literature: An Introduction to the Theory of Genres and Modes*, Clarendon, Oxford.
- Francesconi, S. (2011), 'Images and writing in tourist brochures', *Journal of Tourism and Cultural Change* **9**(4), 341–356.
- Freadman, A. (2012), 'The traps and trappings of genre theory', *Applied Linguistics* **33**(5), 544–563.
- Freedman, K. and Stuhr, P. (2004), Curriculum change for the 21<sup>st</sup> century: Visual culture in art education, in E. W. Eisner and M. D. Day, eds, 'Handbook of Research and Policy in Art Education', Lawrence Erlbaum, Mahwah, NJ, pp. 815–828.
- Frow, J. (2006), *Genre*, Routledge, London.
- Gansner, E. R. and North, S. C. (2000), 'An open graph visualization system and its applications to software engineering', *Software - Practice and Experience* **30**(11), 1203–1233.
- Garrod, B. (2009), 'Understanding the relationship between tourism destination imagery and tourist photography', *Journal of Travel Research* **47**(3), 346–358.
- Gibson, J. J. (1979), *The Ecological Approach to Visual Perception*, Houghton Mifflin, Boston.
- Goffman, E. (1979), *Gender Advertisements*, Macmillan, London.
- Guijarro, A. J. (2011), 'Engaging readers through language and pictures: A case study', *Journal of Pragmatics* **43**(12), 2982–2991.
- Guijarro, A. J. and Hernandez, J. I. (2001), 'Points of departure in news items and tourist brochures: Choices of theme and topic', *Text* **21**(3), 347–371.
- Guijarro, A. J. and Pinar Sanz, M. J. (2008), 'Compositional, interpersonal and representational meanings in a children's narrative: A multimodal discourse analysis', *Journal of Pragmatics* **40**(9), 1601–1619.
- Guo, L. (2004), Multimodality in a biology textbook, in K. L. O'Halloran, ed., 'Multimodal Discourse Analysis: Systemic Functional Perspectives', Continuum, London, pp. 196–219.
- Hallett, R. W. and Kaplan-Weinger, J. (2010), *Official Tourism Websites: A Discourse Analytical Perspective*, Channel View Publications, Bristol.
- Halliday, M. A. K. (1970/2002a), Language structure and language function, in J. J. Webster, ed., 'On Grammar', Vol. 1 of *The Collected Works of M.A.K. Halliday*, Continuum, London, pp. 173–195.
- Halliday, M. A. K. (1973), *Explorations in the Functions of Language*, Arnold, London.
- Halliday, M. A. K. (1975), *Learning How to Mean: Explorations in the Development of Language*, Arnold, London.
- Halliday, M. A. K. (1976a), English system networks, in G. Kress, ed., 'Halliday: System and Function in Language', Oxford University Press, London, pp. 101–135.
- Halliday, M. A. K. (1976b), *Halliday: System and Function in Language*, Oxford University Press, London. Selected papers edited by Günther Kress.



- Halliday, M. A. K. (1978), *Language as a Social Semiotic: The Social Interpretation of Language and Meaning*, Arnold, London.
- Halliday, M. A. K. (1979/2002*b*), Modes of meaning and modes of expression: Types of grammatical structure and their determination by different semantic functions, in J. J. Webster, ed., 'On Grammar', Vol. 1 of *The Collected Works of M.A.K. Halliday*, Continuum, London, pp. 173–195.
- Halliday, M. A. K. (1981), Introduction, in M. A. K. Halliday and J. R. Martin, eds, 'Readings in Systemic Linguistics', Batsford, London, pp. 13–16.
- Halliday, M. A. K. (1985), *Introduction to Functional Grammar*, Arnold, London.
- Halliday, M. A. K. (1987/2003), Language and the order of nature, in J. J. Webster, ed., 'On Language and Linguistics', Vol. 3 of *The Collected Works of M.A.K. Halliday*, Continuum, London, pp. 116–138.
- Halliday, M. A. K. (1992), Some lexicogrammatical features of the Zero Population Growth text, in W. C. Mann and S. A. Thompson, eds, 'Discourse Description: Diverse Linguistic Analyses of a Fund-raising Text', Benjamins, Amsterdam, pp. 327–358.
- Halliday, M. A. K. (1993), On the language of physical science, in M. A. K. Halliday and J. R. Martin, eds, 'Writing Science: Literacy and Discursive Power', Falmer, London, pp. 54–68.
- Halliday, M. A. K. (1994), *Introduction to Functional Grammar*, second edn, Arnold, London.
- Halliday, M. A. K. (2009), *The Essential Halliday*, Continuum, London.
- Halliday, M. A. K. and Fawcett, R. P. (1987), Introduction, in M. A. K. Halliday and R. D. Fawcett, eds, 'New Developments in Systemic Linguistics', Vol. I: Theory and Description, Pinter, London, pp. 1–13.
- Halliday, M. A. K. and Hasan, R. (1986), *Language, Context and Text: Aspects of Language in a Social-Semiotic Perspective*, Deakin University, Victoria.
- Halliday, M. A. K. and Martin, J. R. (1993), *Writing Science: Literacy and Discursive Power*, Falmer, London.
- Halliday, M. A. K. and Matthiessen, C. M. I. M. (1999), *Construing Experience Through Meaning: A Language-based Approach to Cognition*, Continuum, London.
- Halliday, M. A. K. and Matthiessen, C. M. I. M. (2004), *An Introduction to Functional Grammar*, third edn, Arnold, London.
- Hannus, M. and Hyönä, J. (1999), 'Utilization of illustrations during learning of science textbook passages among low- and high-ability children', *Contemporary Educational Psychology* **24**(2), 95–123.
- Hasan, R. (2004), Analysing discursive variation, in L. Young and C. Harrison, eds, 'Systemic Functional Linguistics and Critical Discourse Analysis: Studies in Social Change', Continuum, London, pp. 15–52.
- Hasan, R. (2005), *Language, Society and Consciousness*, Vol. 1 of *The collected works of Ruqaiya Hasan*, Equinox, London.

- Hegarty, M. and Just, M. A. (1993), 'Constructing mental models of machines from text and diagrams', *Journal of Memory and Language* **32**(6), 717–742.
- Held, G. (2005), 'Magazine covers – a multimodal pretext-genre', *Folia Linguistica* **XXXIX**(1-2), 173–196.
- Hendel, R. (1998), *On Book Design*, Yale University Press, New Haven and London.
- Henschel, R. (2003), *GeM Annotation Manual*, second edn, University of Bremen, University of Stirling.
- Henshilwood, C. S. and d'Errico, F. (2011*a*), Middle Stone Age engravings and their significance to the debate on the emergence of symbolic material culture, in C. S. Henshilwood and F. d'Errico, eds, 'Homo Symbolicus: The dawn of language, imagination and spirituality', Benjamins, Amsterdam and Philadelphia, pp. 76–96.
- Henshilwood, C. S. and d'Errico, F., eds (2011*b*), *Homo Symbolicus: The dawn of language, imagination and spirituality*, Benjamins, Amsterdam and Philadelphia.
- Hiippala, T. (2007), Helsinki: A multisemiotic analysis of tourist brochures, Master's thesis, University of Helsinki.
- Hiippala, T. (2012*a*), The localisation of advertising print media as a multimodal process, in W. L. Bowcher, ed., 'Multimodal Texts from Around the World: Linguistic and Cultural Insights', Palgrave, London, pp. 97–122.
- Hiippala, T. (2012*b*), Modelling multimodal genre in print media: a case study of tourist brochures, in P. Couto Cantero, G. Enríquez Veloso, A. Passeri and J. M. Paz Gago, eds, 'Proceedings of the 10th World Congress of the International Association for Semiotic Studies (IASS/AIS)', Universidade da Coruña, A Coruña, pp. 1501–1510.
- Hiippala, T. (2012*c*), 'Reading paths and visual perception in multimodal research, psychology and brain sciences', *Journal of Pragmatics* **44**(3), 315–327.
- Hiippala, T. (2013), 'The interface between rhetoric and layout in multimodal artefacts', *Literary & Linguistic Computing* **28**(3), 461–471.
- Hiippala, T. (in press), Genre analysis, in S. Norris and C. D. Maier, eds, 'Texts, Images and Interactions: A Reader in Multimodality', De Gruyter Mouton, Berlin & New York.
- Hobbs, R. (1998), 'The seven great debates in the media literacy movement', *Journal of Communication* **48**(1), 16–32.
- Hodge, R. and Kress, G. (1988), *Social Semiotics*, Polity, Cambridge.
- Hoey, M. (2001), *Textual Interaction: An Introduction to Written Discourse Analysis*, Routledge, London.
- Holsanova, J. and Holmqvist, K. (2006), 'Entry points and reading paths on newspaper spreads: comparing a semiotic analysis with eye-tracking measurements', *Visual Communication* **5**(1), 65–93.
- Holsanova, J. and Nord, A. (2010), Multimodal design: Media structures, media principles and users meaning-making in printed and digital media, in H.-J. Bucher, T. Gloning and K. Lehnen, eds, 'Neue Medie - neue Formate: Ausdifferenzierung und Konvergenz in der Medienkommunikation', Campus, Frankfurt/New York, pp. 81–103.

- Hopearuoho, A. and Ventola, E. (2009), Multisemiotic marketing and advertising: Globalization versus localization and the media, *in* E. Ventola and A. J. Guijarro, eds, 'The World Told and the World Shown: Multisemiotic Issues', Palgrave, London, pp. 183–204.
- Huemer, B. (2008), Representations of individual and mass: modelling experience through multiple modes in digital art, *in* C. Jones and E. Ventola, eds, 'From Language to Multimodality: New Developments in the Study of Ideational Meaning', Equinox, London, pp. 255–274.
- Hyland, K. (2002), 'Genre: Language, context and literacy', *Annual Review of Applied Linguistics* **22**, 113–135.
- Hyon, S. (1996), 'Genre in three traditions: Implications for ESL', *TESOL Quarterly* **30**(4), 693–721.
- Iedema, R. (2003), 'Multimodality, resemiotization: extending the analysis of discourse as multi-semiotic practice', *Visual Communication* **2**(1), 29–57.
- Ivarsson, J., Linderöth, J. and Säljö, R. (2009), Representations in practices: A socio-cultural approach to multimodality in reasoning, *in* C. Jewitt, ed., 'The Routledge Handbook of Multimodal Analysis', Routledge, London, pp. 201–212.
- Jaworski, A. and Thurlow, C. (2009), Gesture and movement in tourist spaces, *in* C. Jewitt, ed., 'The Routledge Handbook of Multimodal Analysis', Routledge, London, pp. 253–262.
- Jewitt, C. (2009a), Different approaches to multimodality, *in* C. Jewitt, ed., 'The Routledge Handbook of Multimodal Analysis', Routledge, London, pp. 28–39.
- Jewitt, C. (2009b), An introduction to multimodality, *in* C. Jewitt, ed., 'The Routledge Handbook of Multimodal Analysis', Routledge, London, pp. 14–27.
- Jewitt, C. and Bezemer, J. (2010), Multimodal analysis: key issues, *in* L. Litosseliti, ed., 'Research Methods in Linguistics', Continuum, London, pp. 180–197.
- Jewitt, C., ed. (2009c), *The Routledge Handbook of Multimodal Analysis*, Routledge, London.
- Jewitt, C. and Oyama, R. (2001), Visual meaning: a social semiotic approach, *in* T. van Leeuwen and C. Jewitt, eds, 'Handbook of Visual Analysis', SAGE Publications, London, pp. 134–156.
- Jokela, S. (2011), 'Building a facade for Finland: Helsinki in tourism imagery', *The Geographical Review* **101**(1), 53–70.
- Jones, C. and Ventola, E., eds (2008), *From Language to Multimodality: New Developments in the Study of Ideational Meaning*, Equinox, London.
- Kaltenbacher, M. (2004), 'Perspectives on multimodality: from the early beginnings to the state of the art', *Information Design Journal + Document Design* **12**(3), 190–207.
- Kaltenbacher, M. (2006), Culture related linguistic differences in tourist websites: the emotive and the factual: A corpus analysis within the framework of Appraisal, *in* G. Thompson and S. Hunston, eds, 'System and Corpus: Exploring Connections', Equinox, pp. 269–292.

- Kaltenbacher, M. (2007), Systemic functional linguistics and corpus analysis: the language of exaggeration in web-sites of tourism, *in* H. Gruber, M. Kaltenbacher and P. Muntigl, eds, 'Empirical Approaches to Discourse - Empirischeorientierte Ansätze in der Diskursanalyse', Peter Lang, Frankfurt am Main, pp. 89–117.
- Kappas, A. and Olk, B. (2008), 'The concept of visual competence as seen from the perspective of the psychological and brain sciences', *Visual Studies* **23**(2), 162–173.
- Knox, J. (2007), 'Visual-verbal communication on online newspaper home pages', *Visual Communication* **6**(1), 19–53.
- Kok, K. C. A. (2004), Multisemiotic mediation in hypertext, *in* K. L. O'Halloran, ed., 'Multimodal Discourse Analysis: Systemic Functional Perspectives', Continuum, London, pp. 131–159.
- Kong, K. C. C. (2006), 'A taxonomy of the discourse relations between words and visuals', *Information Design Journal* **14**(3), 207–230.
- Kong, K. C. C. (2013), 'A corpus-based study in comparing the multimodality of Chinese- and English-language newspapers', *Visual Communication* **12**(2), 173–196.
- Kostelnick, C. and Hassett, M. (2003), *Shaping Information: The Rhetoric of Visual Conventions*, Southern Illinois University Press, Carbondale.
- Kress, G. (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', Routledge, London & New York, pp. 53–79.
- Kress, G. (2009), What is mode?, *in* C. Jewitt, ed., 'The Routledge Handbook of Multimodal Analysis', Routledge, London, pp. 54–67.
- Kress, G. (2010), *Multimodality: A Social Semiotic Approach to Contemporary Communication*, Routledge, London.
- Kress, G. (2011), 'Partnerships in research: multimodality and ethnography', *Qualitative Research* **11**(3), 239–260.
- Kress, G. and van Leeuwen, T. (1990), *Reading Images*, Deakin University, Deakin.
- Kress, G. and van Leeuwen, T. (1996), *Reading Images: The Grammar of Visual Design*, Routledge, London.
- Kress, G. and van Leeuwen, T. (1998), Front pages: (the critical) analysis of newspaper layout, *in* A. Bell and P. Garrett, eds, 'Approaches to Media Discourse', Blackwell, Oxford, pp. 186–219.
- Kress, G. and van Leeuwen, T. (2001), *Multimodal Discourse: The Modes and Media of Contemporary Communication*, Arnold, London.
- Kress, G. and van Leeuwen, T. (2002), 'Colour as a semiotic mode: notes for a grammar of colour', *Visual Communication* **1**(3), 343–368.
- Kress, G. and van Leeuwen, T. (2006), *Reading Images: The Grammar of Visual Design*, second edn, Routledge, London.
- Kvåle, G. (2010), Invitasjon til destinasjon: Multimodal retorikk i turistkommunikasjon, *in* M. Engebretsen, ed., 'Skrift/bilde/lyd: Analyse av sammensatte tekster', Høyskoleforlaget, Kristiansand, pp. 39–55.

- Labov, W. (2001), *Principles of Linguistic Change*, Vol. 2: Social factors, Blackwell, Oxford.
- Lagerwerf, L., van Hooijdonk, C. M. and Korenberg, A. (2012), 'Processing visual rhetoric in advertisements: Interpretations determined by verbal anchoring and visual structure', *Journal of Pragmatics* **44**(13), 1836–1852.
- Leander, K. M. and Vasudevan, L. (2009), Multimodality and mobile culture, in C. Jewitt, ed., 'The Routledge Handbook of Multimodal Analysis', Routledge, London, pp. 127–139.
- Leckner, S. (2012), 'Presentation factors affecting reading behaviour in readers of newspaper media: an eye-tracking perspective', *Visual Communication* **11**(2), 163–184.
- Leech, G. (1991), The state of the art in corpus linguistics, in K. A. . B. Altenberg, ed., 'English Corpus Linguistics: Studies in Honour of Jan Svartvik', Longman, London, pp. 8–29.
- Lemke, J. L. (1998), Multiplying meaning: Visual and verbal semiotics in scientific text, in J. R. Martin and R. Veel, eds, 'Reading Science: Critical and Functional Perspectives on Discourses of Science', Routledge, New York, NY, pp. 87–113.
- Lemke, J. L. (1999), 'Typology, topology, topography: genre semantics', MS University of Michigan.
- Lemke, J. L. (2000), 'Opening up closure: semiotics across scales', *Annals of the New York Academy of Sciences* **901**(1), 100–111.
- Lemke, J. L. (2005), 'Multimedia genres and traversals', *Folia Linguistica* **XXXIX**(1-2), 45–56.
- Lemke, J. L. (2009), 'Multimodal genres and transmedia traversals: Social semiotics and the political economy of the sign', *Semiotica* **173**(1), 283–297.
- Levon, E. (2010), Organizing and processing your data: The nuts and bolts of quantitative analysis, in L. Litosseliti, ed., 'Research Methods in Linguistics', Continuum, London, pp. 69–90.
- Lim, F. V. (2004a), Developing an integrative multi-semiotic model, in K. O'Halloran, ed., 'Multimodal Discourse Analysis: Systemic Functional Perspectives', Continuum, London, pp. 220–246.
- Lim, F. V. (2004b), Problematizing 'semiotic resource', in E. Ventola, C. Charles and M. Kaltenbacher, eds, 'Perspectives on Multimodality', Benjamins, Amsterdam, pp. 51–64.
- Liu, Y. and O'Halloran, K. L. (2009), 'Intersemiotic texture: analyzing cohesive devices between language and images', *Social Semiotics* **19**(4), 367–388.
- Livingstone, S. (2004), 'Media literacy and the challenge of new information and communication technologies', *The Communication Review* **7**, 3–14.
- Longacre, R. E. (1992), The discourse strategy of an appeals letter, in W. C. Mann and S. A. Thompson, eds, 'Discourse Description: Diverse Linguistic Analyses of a Fund-raising Text', Benjamins, Amsterdam, pp. 109–130.
- Luke, C. (2003), 'Pedagogy, connectivity, multimodality, and interdisciplinarity', *Reading Research Quarterly* **38**(3), 397–403.

- Lukin, A., Moore, A., Herke, M., Wegener, R. and Wu, C. (2008), 'Halliday's model of register revisited and explored', *Linguistics and the Human Sciences* **4**(2), 187–243.
- MacCannell, D. (1989), *The Tourist: A New Theory of the Leisure Class*, revised edn, Schocken, New York.
- MacEachren, A. M. (1995), *How Maps Work: Representation, Visualization and Design*, Guilford, New York.
- Machin, D. (2004), 'Building the world's visual language: The increasing global importance of image banks in corporate media', *Visual Communication* **3**(3), 316–336.
- Machin, D. (2009), Multimodality and theories of the visual, in C. Jewitt, ed., 'The Routledge Handbook of Multimodal Analysis', Routledge, London, pp. 181–190.
- Machin, D. and van Leeuwen, T. (2005), 'Language style and lifestyle: the case of a global magazine', *Media, Culture and Society* **27**(4), 577–600.
- Maiorani, A. (2008), Movies 'reloaded' into commercial reality: representational structures in *The Matrix* trilogy promotional posters, in C. Jones and E. Ventola, eds, 'From Language to Multimodality: New Developments in the Study of Ideational Meaning', Equinox, London, pp. 275–296.
- Maiorani, A. (2009), Developing the metafunctional framework for analysing multimodal hypertextual identity construction, in E. Ventola and A. J. Guijarro, eds, 'The World Told and the World Shown: Multisemiotic Issues', Palgrave, London, pp. 220–241.
- Mann, W. C. (2005), 'Rhetorical Structure Theory website'.  
**URL:** <http://www.sfu.ca/rst/>
- Mann, W. C. and Matthiessen, C. M. (1990), Functions of language in two frameworks, Technical report. ISI/RR-90-290, Information Sciences Institute (ISI), Los Angeles, CA.
- Mann, W. C., Matthiessen, C. M. and Thompson, S. A. (1992), Rhetorical structure theory and text analysis, in W. C. Mann and S. A. Thompson, eds, 'Discourse Description: Diverse Linguistic Analyses of a Fund-raising Text', Benjamins, Amsterdam, pp. 39–78.
- Mann, W. C. and Thompson, S. A. (1987), Rhetorical Structure Theory: A theory of text organization, Technical report. ISI/RS-87-190, University of Southern California, Information Sciences Institute (ISI).
- Mann, W. C. and Thompson, S. A. (1988), 'Rhetorical Structure Theory: Toward a functional theory of text organization', *Text* **8**(3), 243–281.
- Manovich, L. (2001), *Language of the New Media*, second edn, MIT Press, Cambridge, MA.
- Manovich, L. (2012), How to compare one million images?, in D. M. Berry, ed., 'Understanding Digital Humanities', Palgrave, Basingstoke, pp. 249–278.
- Martin, J. R. (1991), 'Intrinsic functionality: implications for contextual theory', *Social Semiotics* **1**(1), 99–162.
- Martin, J. R. (1992), *English Text: System and Structure*, Benjamins, Amsterdam.
- Martin, J. R. (1994), 'Modelling big texts: A systemic functional approach to multi-genericity', *Network* **21**, 29–52.

- Martin, J. R. (1996), Types of structure: Deconstructing notions of constituency in clause and text, *in* E. H. Howy and D. R. Scott, eds, 'Computation and Conversational Discourse: Burning Issues - an Interdisciplinary Account', Springer, Heidelberg, pp. 39–66.
- Martin, J. R. (1997), Analysing genre: functional parameters, *in* F. Christie and J. R. Martin, eds, 'Genre and Institutions: Social Processes in the Workplace and School', Continuum, London, pp. 3–39.
- Martin, J. R. (1999), Modelling context: A crooked path of progress in contextual linguistics, *in* M. Ghadessy, ed., 'Text and Context in Functional Linguistics', Vol. 169 of *Amsterdam Studies in the Theory and History of Linguistic Science: Series IV – Current Issues In Linguistic Theory*, Benjamins, Amsterdam, pp. 25–61.
- Martin, J. R. (2009), Discourse studies, *in* M. A. K. Halliday and J. J. Webster, eds, 'Continuum Companion to Systemic Functional Linguistics', Continuum, London, pp. 154–165.
- Martin, J. R. (2011), Systemic functional linguistics, *in* K. Hyland and B. Paltridge, eds, 'Continuum Companion to Discourse Analysis', Continuum, London, pp. 101–119.
- Martin, J. R. (2012), Writing and genre studies, *in* C. A. Chapelle, ed., 'The Encyclopedia of Applied Linguistics', Wiley-Blackwell, Oxford, pp. 6258–6266.
- Martin, J. R. and Matthiessen, C. M. (1991), Systemic typology and topology, *in* F. Christie, ed., 'Literacy in Social Processes. Papers from the Inaugural Australian Systemic Functional Linguistics Conference, held at Deakin University, January 1990', Centre for Studies of Language in Education, Northern Territory University, Darwin, pp. 345–383.
- Martin, J. R. and Rose, D. (2008), *Genre Relations: Mapping Culture*, Equinox, London.
- Martin, J. R. and Veel, R. (1998), *Reading Science: Critical and Functional Perspectives on Discourses of Science*, Routledge, New York, NY.
- Martinec, R. (2003), 'The social semiotics of text and image in Japanese and English software manuals and other procedures', *Social Semiotics* **13**(1), 43–69.
- Martinec, R. (2005), Topics in multimodality, *in* R. Hasan, C. M. Matthiessen and J. J. Webster, eds, 'Continuing Discourse on Language: A Functional Perspective', Vol. 1, pp. 157–181.
- Martinec, R. (2013), 'Nascent and mature uses of a semiotic system: the case of image-text relations', *Visual Communication* **12**(2), 147–172.
- Martinec, R. and Salway, A. (2005), 'A system for image-text relations in new (and old) media', *Visual Communication* **4**(3), 337–371.
- Martínez-Camino, G. and Pérez-Saiz, M. (2012), 'A pragmatics theory on television advertising', *Journal of Pragmatics* **44**(4), 453–473.
- Martínez Lirola, M. and Chovanec, J. (2012), 'The dream of a perfect body come true: Multimodality in cosmetic surgery advertising', *Discourse & Society* **23**(5), 487–507.
- Matthiessen, C. M. (2007), The multimodal page: A systemic functional exploration, *in* W. L. Bowcher and T. D. Royce, eds, 'New Directions in the Analysis of Multimodal Discourse', Lawrence Erlbaum, Mahwah, NJ, pp. 1–62.

- Mayer, R. E. (2005), Principles for reducing extraneous processing in multimedia learning: Coherence, signaling, redundancy, spatial contiguity, and temporal contiguity principles, *in* R. E. Mayer, ed., ‘The Cambridge Handbook of Multimedia Learning’, Cambridge University Press, Cambridge, pp. 183–200.
- Mayer, R. E. and Moreno, R. (2003), ‘Nine ways to reduce cognitive load in multimedia learning’, *Educational Psychologist* **38**(1), 43–52.
- Messaris, P. (1994), *Visual “Literacy”: Image, Mind & Reality*, Westview, Boulder, CO.
- Michel, J.-B., Shen, Y. K., Aiden, A. P., Veres, A., Gray, M. K., Brockman, W., The Google Books Team, Pickett, J. P., Hoiberg, D., Clancy, D., Norvig, P., Orwant, J., Pinker, S., Nowak, M. A. and Aiden, E. L. (2011), ‘Quantitative analysis of culture using millions of digitized books’, *Science* **331**(6014), 176–182.
- Miller, C. R. (1984), ‘Genre as social action’, *Quarterly Journal of Speech* **70**, 151–167.
- Mitchell, T. F. (1957), The language of buying and selling in cyrenaica: a situational statement, *in* ‘Principles of Firthian Linguistics’, Longman, London, pp. 167–200. Reprinted in 1975.
- Mitchell, W. J. T. (2005), ‘There are no visual media’, *Journal of Visual Culture* **4**(2), 257–266.
- Mitocariu, E., Anechitei, D. A. and Cristea, D. (2013), Comparing discourse tree structures, *in* A. Gelbukh, ed., ‘Computational Linguistics and Intelligent Text Processing’, Vol. 7816 of *Lecture Notes in Computer Science*, Springer, Berlin & Heidelberg, pp. 513–522.
- Molina, A. and Esteban, A. (2006), ‘Tourism brochures: Usefulness and image’, *Annals of Tourism Research* **33**(4), 1036–1056.
- Monaghan, J. (1979), *The Neo-Firthian Tradition and its Contribution to General Linguistics*, Niemeyer, Tübingen.
- Motta-Roth, D. and Nascimento, F. S. (2009), ‘Transitivity in visual grammar: concepts and applications’, *Revista Linguagem Ensino* **12**(2), 319–349.
- Norris, S. (2004), *Analyzing Multimodal Interaction: A Methodological Framework*, Routledge, London.
- Norris, S. (2012), Multimodal communication: Overview, *in* C. A. Chapelle, ed., ‘The Encyclopedia of Applied Linguistics’, Wiley-Blackwell, Oxford, pp. 3977–3978.
- Norris, S. and Jones, R. H. (2005), *Discourse in Action: Introducing Mediated Discourse Analysis*, Routledge, London.
- O’Halloran, K. L. (1999a), ‘Interdependence, interaction and metaphor in multisemiotic texts’, *Social Semiotics* **9**(3), 317–354.
- O’Halloran, K. L. (1999b), ‘Towards a systemic functional analysis of multisemiotic mathematics texts’, *Semiotica* **124**(1-2), 1–30.
- O’Halloran, K. L. (2004a), Visual semiosis in film, *in* K. L. O’Halloran, ed., ‘Multimodal Discourse Analysis: Systemic Functional Perspectives’, Continuum, London, pp. 109–130.
- O’Halloran, K. L. (2005), *Mathematical Discourse: Language, Symbolism and Visual Images*, Continuum, London.



- O'Halloran, K. L. (2008a), Inter-semiotic expansion of experiential meaning: hierarchical scales and metaphor in mathematics discourse, in C. Jones and E. Ventola, eds, 'From Language to Multimodality: New Developments in the Study of Ideational Meaning', Equinox, London, pp. 231–254.
- O'Halloran, K. L. (2008b), 'Systemic functional-multimodal discourse analysis (SFMDA): Constructing ideational meaning using language and visual imagery', *Visual Communication* 7(4), 443–475.
- O'Halloran, K. L. (2009a), Historical changes in the semiotic landscape: from calculation to computation, in C. Jewitt, ed., 'The Routledge Handbook of Multimodal Analysis', Routledge, London, pp. 98–113.
- O'Halloran, K. L. (2009b), Multimodal analysis and digital technology, in 'Interdisciplinary Perspectives on Multimodality: Theory and Practice. Proceedings of the Third International Conference on Multimodality', Palladino, Campobasso.
- O'Halloran, K. L. (2011), Multimodal discourse analysis, in K. Hyland and B. Paltridge, eds, 'Continuum Companion to Discourse Analysis', Continuum, London, pp. 120–137.
- O'Halloran, K. L., ed. (2004b), *Multimodal Discourse Analysis: Systemic Functional Perspectives*, Continuum, London.
- O'Halloran, K. L. and Lim, F. V. (2009), Sequential visual discourse frames, in E. Ventola and A. J. Guijarro, eds, 'The World Told and the World Shown: Multisemiotic Issues', Palgrave, London, pp. 139–156.
- O'Halloran, K. L., Podlasov, A., Chua, A. and E, Marissa K. L. (2012), 'Interactive software for multimodal analysis', *Visual Communication* 11(3), 363–381.
- O'Halloran, K. L. and Smith, B. A. (2012), Multimodality and technology, in C. A. Chapelle, ed., 'The Encyclopedia of Applied Linguistics', Wiley-Blackwell, Oxford, pp. 4089–4094.
- O'Halloran, K. L., Tan, S., Smith, B. A. and Podlasov, A. (2010), 'Challenges in designing digital interfaces for the study of multimodal phenomena', *Information Design Journal* 18(1), 2–21.
- O'Halloran, K. L., Tan, S., Smith, B. A. and Podlasov, A. (2011), 'Multimodal analysis within an interactive software environment: critical discourse perspectives', *Critical Discourse Studies* 8(2), 109–125.
- Okun, O., Doermann, D. and Pietikäinen, M. (1999), Page segmentation and zone classification: the state of the art, Technical report. LAMP-TR-036, University of Maryland, Language and Media Processing Laboratory.
- Orlikowski, W. J. and Yates, J. (1994), 'Genre repertoire: The structuring of communicative practices in organizations', *Administrative Science Quarterly* 39, 541–574.
- O'Toole, M. (1994), *The Language of Displayed Art*, Leicester University Press, London.
- O'Toole, M. (2004), Opera Ludentes: the Sydney Opera House at work and play, in K. L. O'Halloran, ed., 'Multimodal Discourse Analysis: Systemic Functional Perspectives', Continuum, London, pp. 11–27.
- O'Toole, M. (2011), *The Language of Displayed Art*, second edn, Routledge, New York.

- Pagani, G. (2009), Citizenship and semiotics: Towards a multimodal analysis of representations of the relationship between the state and the citizen, *in* E. Ventola and A. J. Guijarro, eds, 'The World Told and the World Shown: Multisemiotic Issues', Palgrave, London, pp. 90–103.
- Paltridge, B. (1997), *Genre, Frames and Writing in Research Settings*, Benjamins, Amsterdam.
- Paltridge, B. (2002), 'Thesis and dissertation writing: an examination of published advice and actual practice', *English for Specific Purposes* **21**, 125–143.
- Parodi, G. (2010), 'Research challenges for corpus cross-linguistics and multimodal texts', *Information Design Journal* **18**(1), 69–73.
- Phillips, B. J. and McQuarrie, E. F. (2002), 'The development, change, and transformation of rhetorical style in magazine advertisements 1954-1999', *Advertising and Society Review* **31**(4), 1–13.
- Piazza, R., Bednarek, M. and Rossi, F. (2011), *Telecinematic Discourse: Approaches to the Language of Films and Television Series*, Benjamins, Amsterdam.
- Podlasov, A., Tan, S. and O'Halloran, K. L. (2012), 'Interactive state-transition diagrams for visualization of multimodal annotation', *Intelligent Data Analysis* **16**(4), 683–702.
- Rangoni, Y., Belaïd, A. and Vajda, S. (2012), 'Labelling logical structures of document images using a dynamic perceptive neural network', *International Journal on Document Analysis and Recognition* **15**(1), 45–55.
- Reichenberger, K., Rondhuis, K., Kleinz, J. and Bateman, J. A. (1996), Effective presentation of information through page layout: a linguistically-based approach, Technical Report Arbeitspapiere der GMD 970, Institut für Integrierte Publikations- und Informationssysteme (IPSI), GMD, Darmstadt.  
**URL:** <http://www.cs.uic.edu/~ifc/mmwsproc/reichen/page-layout.html>
- Riley, H. (2004), 'Perceptual modes, semiotic codes, social mores: a contribution towards a social semiotics of drawing', *Visual Communication* **3**(3), 294–315.
- Rose, D. (2006), 'A systemic functional approach to language evolution', *Cambridge Archeological Journal* **16**, 73–96.
- Rose, D. (2011), Genre in the Sydney School, *in* J. P. Gee and M. Handford, eds, 'The Routledge Handbook of Discourse Analysis', Routledge, London.
- Royce, T. D. (1998), 'Synergy on the page: Exploring intersemiotic complementarity in page-based multimodal text', *JASFL Occasional Papers* **1**(1), 25–49.
- Royce, T. D. (2007), Intersemiotic complementarity: A framework for multimodal discourse analysis, *in* T. D. Royce and W. L. Bowcher, eds, 'New Directions in the Analysis of Multimodal Discourse', Lawrence Erlbaum, Mahwah, NJ, pp. 63–109.
- Scarles, C. (2004), 'Mediating landscapes: The processes and practices of image construction in tourist brochures of Scotland', *Tourist Studies* **4**(1), 43–67.
- Scolari, C. A. (2009), 'Transmedia storytelling: Implicit consumers, narrative worlds, and branding in contemporary media production', *International Journal of Communication* **3**, 586–606.

- Scollon, R. and Scollon, S. W. (2009), Multimodality and language: A retrospective and prospective view, in C. Jewitt, ed., 'The Routledge Handbook of Multimodal Analysis', Routledge, London, pp. 170–180.
- Scott, M. (2010), 'Book Review: Multimodality and Genre: A Foundation for the Systematic Analysis of Multimodal Documents', *Visual Communication* 9(2), 241–245.
- Seppänen, J. (2005), *Visuaalinen kulttuuri: Teoriaa ja metodeja mediakuvan tulkitsijalle*, Vastapaino, Tampere. English title: 'Visual culture: Theory and methods for analysing representations in media'.
- Siirtola, H., Rähkä, K.-J., Säily, T. and Nevalainen, T. (2010), Information visualization for corpus linguistics: Towards interactive tools, in S. Liu, M. X. Zhou, G. Carenini and H. Qu, eds, 'Proceedings of the First International Workshop on Intelligent Visual Interfaces for Text Analysis', ACM, New York, pp. 33–36.
- Smith, B. A., Tan, S., Podlasov, A. and O'Halloran, K. L. (2011), 'Analyzing multimodality in an interactive digital environment: Software as metasemiotic tool', *Social Semiotics* 21(3), 369–380.
- Stede, M. (2008), 'Disambiguating rhetorical structure', *Research on Language & Computation* 6(3-4), 311–332.
- Stenglin, M. and Djonov, E. (2010), Unpacking narrative in a hypermedia 'artedventure' for children, in C. R. Hoffmann, ed., 'Narrative Revisited: Telling a story in the age of new media', Vol. 199 of *Pragmatics and Beyond*, John Benjamins, Amsterdam, pp. 185–212.
- Stöckl, H. (2004), In between modes: language and image in printed media, in E. Ventola, C. Charles and M. Kaltenbacher, eds, 'Perspectives on Multimodality', Benjamins, Amsterdam, pp. 9–30.
- Stöckl, H. (2005), 'Typography: body and dress of a text — a signing mode between language and image', *Visual Communication* 4(2), 210–214.
- Swales, J. M. (1990), *Genre analysis: English in Academic and Research Settings*, Cambridge University Press, Cambridge.
- Taboada, M. and Habel, C. (2013), 'Rhetorical relations in multimodal documents', *Discourse Studies* 15(1), 65–89.
- Taboada, M. and Mann, W. C. (2006a), 'Applications of rhetorical structure theory', *Discourse Studies* 8(4), 567–588.
- Taboada, M. and Mann, W. C. (2006b), 'Rhetorical structure theory: looking back and moving ahead', *Discourse Studies* 8(3), 423–459.
- Tan, S. (2009), A systemic functional framework for the analysis of corporate television advertisements, in E. Ventola and A. J. Guijarro, eds, 'The World Told and the World Shown: Multisemiotic Issues', Palgrave, London, pp. 157–182.
- Tan, S. (2010), 'Modelling engagement in a web-based advertising campaign', *Visual Communication* 9(1), 91–115.
- Teich, E. (2009), Linguistic computing, in M. A. K. Halliday and J. J. Webster, eds, 'Continuum Companion to Systemic Functional Linguistics', Continuum, London, pp. 113–127.

- Thibault, P. J. (1987), An interview with Michael Halliday, *in* R. Steele and T. Threadgold, eds, 'Language Topics: Essays in honour of Michael Halliday', Vol. 2, Benjamins, Amsterdam, pp. 601–627.
- Thibault, P. J. (2000), The multimodal transcription of a television advertisement: Theory and practice, *in* A. Baldry, ed., 'Multimodality and Multimediality in the Distance Learning Age', Palladino, Campobasso, pp. 311–385.
- Thomas, M. (2007), Querying multimodal annotation: A concordancer for GeM, *in* 'Proceedings of the Linguistic Annotation Workshop', Association for Computational Linguistics, Prague, Czech Republic, pp. 57–60.  
**URL:** <http://www.aclweb.org/anthology/W/W07/W07-1510>
- Thomas, M. (2009a), Developing multimodal texture, *in* E. Ventola and A. J. Guijarro, eds, 'The World Told and the World Shown: Multisemiotic Issues', Palgrave, London, pp. 39–55.
- Thomas, M. (2009b), Localizing pack messages: A framework for corpus-based cross-cultural multimodal analysis, PhD thesis, University of Leeds.
- Thompson, G. (2004), *Introducing Functional Grammar*, second edn, Arnold, London.
- Thurlow, C. and Jaworski, A. (2003), 'Communicating a global reach: Inflight magazines as a globalizing genre in tourism', *Journal of Sociolinguistics* 7(4), 579–606.
- Thurlow, C. and Jaworski, A. (2010), *Tourism Discourse: Language and Global Mobility*, Palgrave, Basingstoke.
- Toms, E. G. and Campbell, D. G. (1999), Genre as interface metaphor: Exploiting form and function in digital environments, *in* 'Proceedings of the 32nd Hawaii International Conference on System Sciences'.
- Tseng, C. (2008), Cohesive harmony in filmic text, *in* L. Unsworth, ed., 'Multimodal Semiotics: Functional Analysis in Contexts of Education', Continuum, London, pp. 87–104.
- Tseng, C. (2009), Cohesion in Film, and the construction of filmic thematic configurations: A Functional Perspective, PhD thesis, Faculty of Linguistics and Literary Sciences, University of Bremen.
- Tseng, C. and Bateman, J. A. (2010), Chain and choice in filmic narrative: an analysis of multimodal narrative construction in *The Fountain*, *in* C. R. Hoffmann, ed., 'Narrative Revisited: Telling a story in the age of new media', Vol. 199 of *Pragmatics and Beyond*, John Benjamins, Amsterdam, pp. 213–244.
- Tseng, C. and Bateman, J. A. (2012), 'Multimodal narrative construction in Christopher Nolan's *Memento*: a description of analytic method', *Visual Communication* 11(1), 91–119.
- Twyman, M. (1979), A schema for the study of graphic language, *in* P. A. Kolers, M. E. Wrolstad and H. Bouma, eds, 'Processing of Visible Language', Vol. 1, Plenum Press, New York, pp. 117–150.
- Twyman, M. (1986), Articulating graphic language: A historical perspective, *in* M. E. Wrolstad and D. F. Fisher, eds, 'Toward a New Understanding of Literacy', Praeger, New York, pp. 188–251.

- Valdeón, R. A. (2009), 'Info-promotional material discourse and its translation: The case of the Asturian tourist board texts', *Across Languages and Cultures* **10**(1), 21–47.
- van Leeuwen, T. (1991), 'Conjunctive structure in documentary film and television', *Continuum* **5**(1), 76–114.
- van Leeuwen, T. (1993), 'Genre and field in critical discourse analysis: a synopsis', *Discourse & Society* **4**(2), 193–223.
- van Leeuwen, T. (1996), Moving English: The visual language of film, in S. Goodman and D. Graddol, eds, 'Redesigning English: New Texts, New Identities', Routledge, pp. 81–105.
- van Leeuwen, T. (1999), *Speech, Music, Sound*, Macmillan, Basingstoke.
- van Leeuwen, T. (2005a), *Introducing Social Semiotics*, Routledge, London.
- van Leeuwen, T. (2005b), Multimodality, genre and design, in S. Norris and R. H. Jones, eds, 'Discourse in Action: Introducing Mediated Discourse Analysis', Routledge, London, pp. 73–94.
- van Leeuwen, T. (2005c), Three models of interdisciplinarity, in R. Wodak and P. A. Chilton, eds, 'A New Agenda in (Critical) Discourse Analysis: Theory, Methodology and Interdisciplinarity', Benjamins, Amsterdam, pp. 3–18.
- van Leeuwen, T. (2005d), 'Typographic meaning', *Visual Communication* **4**(2), 137–143.
- van Leeuwen, T. (2006), 'Towards a semiotics of typography', *Information Design Journal and Document Design* **14**(2), 139–155.
- van Leeuwen, T. (2011), *The Language of Colour: An Introduction*, Routledge, London.
- Van Mulken, M., Le Pair, R. and Forceville, C. (2010), 'The impact of perceived complexity, deviation and comprehension on the appreciation of visual metaphor in advertising across three European countries', *Journal of Pragmatics* **42**(12), 3418–3430.
- Velasco, J. A. P. (2012), 'A corpus-based approach to the multimodal analysis of specialized knowledge', *Language Resources and Evaluation* .
- Ventola, E. (1987), *The Structure of Social Interaction: A Systemic Approach to the Semiotics of Service Encounters*, Pinter, London.
- Ventola, E. (2005), Multimodality, multimediality and multiliteracies: A new era for the press and using the press, in H. E. H. Lenk and A. Chesterman, eds, 'Presstextsorten im Vergleich — Contrasting Text Types in the Press', Olms, Hildesheim, pp. 351–387.
- Ventola, E. (2011), Semiotisation processes of space: From drawing our homes to styling them, in K. L. O'Halloran and B. A. Smith, eds, 'Multimodal Studies: Multiple Approaches and Domains', Routledge, London, pp. 220–238.
- Ventola, E., Charles, C. and Kaltenbacher, M., eds (2004), *Perspectives on Multimodality*, Benjamins, Amsterdam.
- Ventola, E. and Guijarro, A. J., eds (2009), *The World Told and the World Shown: Multisemiotic Issues*, Palgrave, London.
- Vertesi, J. (2008), 'Mind the gap: The London underground map and users' representations of urban space', *Social Studies of Science* **38**(1), 7–33.
- Waller, R. H. W. (1984), 'Tourist guides – a consumer test', *Information Design Journal* **4**(2), 154–160.

- Waller, R. H. W. (1987), The typographic contribution to language, PhD thesis, Department of Typography and Graphic Communication, University of Reading.
- Waller, R. H. W. (1990), Typography and discourse, in R. Barr, M. L. Kamil, P. Mosenthal and P. D. Pearson, eds, 'Handbook of Reading Research', Vol. II, Lawrence Erlbaum, Mahwah, NJ, pp. 341–380.
- Waller, R. H. W. (2012), 'Graphic literacies for a digital age: The survival of layout', *The Information Society* **28**(4), 236–252.
- Waller, R. H. W., Delin, J. and Thomas, M. (2012), 'Towards a pattern language approach to document description', *Discours* **10**.
- Walmsley, P. (2007), *XQuery*, O'Reilly, Sebastopol, CA.
- Warner, C. (2009), *Media Selling: Television, Print, Internet, Radio*, fourth edn, Wiley-Blackwell, Oxford.
- White, P. (2010), 'Grabbing attention: the importance of modal density in advertising', *Visual Communication* **9**(4), 371–397.
- White, P. (2012), 'Multimodality's challenge to marketing theory: A discussion', *Multimodal Communication* **1**(3), 305–323.
- Wildfeuer, J. (2012), 'Intersemiosis in film: Towards a new organisation of semiotic resources in multimodal filmic text', *Multimodal Communication* **1**(3), 277–304.
- Williamson, J. H. (1986), 'The grid: History, use, and meaning', *Design Issues* **3**(2), 15–30.
- World Tourism Organization (2012), *UNWTO Tourism Highlights, 2012 Edition*, World Tourism Organization (UNWTO), Madrid.
- Yang, D. F. (2008), Representing experience: the co-articulation of verbiage and image in multimodal text, in C. Jones and E. Ventola, eds, 'From Language to Multimodality: New Developments in the Study of Ideational Meaning', Equinox, London, pp. 297–312.
- Yui Ling Ip, J. (2008), 'Analyzing tourism discourse: A case study of a Hong Kong travel brochure', *LCOM Papers* **1**, 1–19.
- Zappavigna, M. (2010), Visualizing logogenesis: Preserving the dynamics of meaning, in S. Dreyfus, S. Hood and M. Stenglin, eds, 'Semiotic Margins: Meaning in Multimodalities', Continuum, London, pp. 211–228.
- Zhao, S. (2010a), Intersemiotic relations as logogenetic patterns: Towards the restoration of the time dimension in hypertext description, in M. Bednarek and J. R. Martin, eds, 'New Discourse on Language: Functional Perspectives on Multimodality, Identity, and Affiliation', Continuum, London, pp. 195–218.
- Zhao, S. (2010b), Rank in visual grammar: Some implications for multimodal discourse analysis, in A. Mahboob and N. K. Knight, eds, 'Applicable Linguistics', Continuum, London, pp. 251–266.