Interpersonal Meaning in Textbooks for Teaching English as a Foreign Language in China: A Multimodal Approach

CHEN Yumin

PhD

University of Sydney

2009
Interpersonal Meaning in Textbooks for Teaching English as a Foreign Language in China: A Multimodal Approach

CHEN Yumin

A thesis submitted in fulfillment of the requirements for the degree of Doctor of Philosophy

Department of Linguistics University of Sydney

April 2009
Declaration

I certify that this thesis is my original work towards the Degree of Doctor of Philosophy. To the best of my knowledge, due reference has been made to all information sources and literature used in this thesis, as well as to any material previously published or written by another person. This thesis does not incorporate without appropriate acknowledgments any material previously submitted for another degree.

Signature of Candidate

_______________________________________
Date

_______________________________________
Abstract

There is increasing awareness among linguists that discourse analysis inevitably involves analyses of meanings arising from the combination of multiple modes of communication. The evolving multimodal pedagogic environment for teaching English as a foreign language (henceforth EFL), among other communicative contexts, calls for a social, semiotic, and linguistic explanation. Situated within the theoretical landscape of social semiotics and in the pedagogic context of EFL education, the present study aims to elucidate how linguistic and visual semiotic resources are co-deployed to construe interpersonal meaning in multimodal textbooks.

The data drawn upon are eighteen EFL textbooks for primary and secondary schooling, published by People’s Education Press between 2002 and 2006. The research design consists of three complementary sub-studies. First, it investigates the ways in which the semantic regions of ENGAGEMENT and GRADUATION can be modelled in multimodal texts, with special reference to the interplay of voices in textbook discourse. The second sub-study analyzes how verbal and visual semiotic resources are co-deployed to construe the ‘emotion and attitude’ goal highlighted in curriculum standards, with a particular focus on verbiage-image relations. Third, it extends the linguistic concept ‘modality’ to multimodal discourse, exploring coding orientation in texts for different educational contexts and between different constituent genres.

The main findings of this thesis are as follows: (1) A range of multimodal resources (i.e. labelling, dialogue balloon, jointly-constructed text, illustration and highlighting) are identified as enabling editor voice to negotiate meanings with reader voice and character voice. It is found that the way in which an ENGAGEMENT value can be scaled is strongly associated with the intrinsic property of the given multimodal resource. The interaction between multiple voices is closely related to contact, social distance, and point of view. (2) It is shown that images play an essential role in realizing attitudinal meanings. Together with verbal APPRAISAL
resources, visual semiotic features work to position the readers in ways that align them to set pedagogic goals, guiding them in completing jointly-constructed texts. Moreover, an attitudinal shift from an emotional release to a more institutionalized type of evaluation can be identified as students advance through the school years. (3) It is argued that what counts as real in multimodal texts is socially defined and specific to a given communicative context. The nature of pedagogic discourse should be taken into account when visual displays are produced for pedagogic materials.

The implications of this study include both theoretical and pedagogic aspects. Theoretically it adapts and extends APPRAISAL analysis to multimodal discourse, exploring the intersemiotic complementarity and co-instantiation in construing global evaluative stance. This semiotic exploration, in return, suggests ways in which discourse analysis may help textbook users better understand and interpret the multimodal features. With the affordances as well as limitations of semiotic resources made explicit, we may have one step further towards a comprehensive and critical understanding of multimodal construal of interpersonal meaning in pedagogic materials.

**Key words:** interpersonal meaning, multimodal discourse analysis, social semiotics, APPRAISAL, textbooks for teaching English as a foreign language
摘 要

“多模态语篇”指包含一种以上符号系统（如文本、图像、声音、动作等）并通过多种符号资源内部的互动来实现意义的复合语篇。在近年来兴起的多模态语篇研究中，语言不再被视为孤立的研究对象，而更多地与参加意义构建的其他符号资源相联系进行分析和解释。教育语境中的多模态现象日益引起语言学界的关注，本论文从语篇分析的角度出发，运用系统功能语言学理论，尤其是其中的评价系统和社会符号学视觉分析法，以人民教育出版社2002年至2006年出版的中小学外语教科书为例，探讨教科书语篇中的语言和图像符号构建人际意义的体现形式和相互关系。此外，基于语言学范畴的人际意义研究模式如何在多模态语篇中发展和应用，也是本论文关注的理论重点。

本论文探讨的主要问题包括：（1）考察外语教科书语篇的多声性质，分析多模态的介入和分级资源在语篇多声互动中的作用；（2）探讨语言和图像符号系统在构建语篇态度意义中的相互关系，及其对实现当前外语教育中“情感态度”目标的意义；（3）研究不同学习阶段多模态外语教科书语篇中多样化的情态编码导向，并讨论如何在教育语境中适当运用表达不同情态意义的图像。

研究结果表明：（1）外语教科书语篇中存在编写者、读者和故事角色三种声音，运用评价系统中的介入子系统对语篇多声互动的实现方式进行分析，得出介入意义在多模态语篇中的具体体现形式包括单词标签、对话泡、共建语篇、插图和图像中高亮的文字等，这些多模态符号资源是教科书编写者联盟读者的重要手段，分别表达否认、声明、引发和摘引四种介入意义。从分级的角度来看，这些符号资源能够实现不同程度的介入意义，取决于故事角色在指示读者方面所承担的责任的多寡、共建语篇的典型性、以及插图在联系文本各部分中的作用等因素。此外，研究还根据视觉交际中的接触、社会距离和视角等参数讨论了语篇多声互动的特征。（2）语言和图像符号资源在构建语篇整体态度取向时存在互补和同示两种关系，通过运用评价系统中的态度子系统对不同阶段的教科书进行分析，可以看出态度意义（包括情感、判定和鉴别）在图像和语言中的具体实现方式包括内嵌式和激发式。从个体发生的角度来审视课程标准中的情感态度目标，可以归纳出其实现过程是从个体情感释放到评价群体行为和社会现象的渐进积累过程。
（3）运用社会符号学关于情态编码导向的理论，分析不同阶段教科书关于同一主题的不同图像类型，可以发现视觉情态的选择与交际双方的认同（solidarity）存在相互依赖的关系。此外，对同一宏观语篇中不同成分语篇的分析表明，除了语旨以外，语域也是影响视觉情态选择的另一重要因素。研究还尝试从社会学理论中关于教育语篇性质的论述出发，讨论感官编码导向在教育语篇中的比例问题。

本研究的创新之处及意义在于：（1）在理论上，论文扩展了评价系统在多模态语篇中的应用，探讨了以往多模态研究中较少涉及的介入和分级子系统。研究表明，在多模态语篇中，同一种介入意义可以由不同的符号资源来体现，同一种符号资源在不同的语境下可体现不同的介入意义，表达同一种介入意义的同种符号资源又可体现不同强度的赋值语义。（2）本研究的另一理论贡献在于，考察了实现态度意义的符际互补和同示关系，并在多模态交际中探讨了系统功能语言学新近发展中提出的承担（commitment）和耦合（coupling）两个概念。（3）论文尝试从社会符号学角度阐释当前课程标准中强调的“情感态度”目标，为该目标的实现过程提供可靠的语言学和符号学描述。（4）论文将视觉分析中的情态编码导向与教育语篇的性质相联系，指出在教育语境中运用图像需要注意的一些问题。

综上所述，本研究表明语言和图像等多模态符号资源在实现教科书语篇人际意义方面发挥着重要作用，对语言学人际意义研究和教科书的编写设计具有一定的理论价值和现实意义。

关键词：
人际意义，多模态语篇分析，社会符号学，评价系统，外语教科书
Acknowledgements

This thesis is the outcome of a cotutelle PhD programme between The University of Sydney and Sun Yat-sen University, supervised by Professor J. R. Martin and Professor Huang Guowen. First and foremost, I am immensely grateful to my supervisor Professor J.R. Martin, who has always been so generous in making time to answer my questions and discuss my work. I am greatly indebted to him for his detailed and insightful comments on the drafts of my thesis, his inspirational guidance, and the hours of discussion, which have continuously challenged my thinking and encouraged me to push the boundaries of my work. My heartfelt gratitude is also due to Professor Huang Guowen, my supervisor in Sun Yat-sen University. His invaluable support and constant encouragement have been crucial in guiding me through various stages from the very first day of my linguistic journey to the last chapter of this thesis. The loyalty I have developed to this academic father has evolved into my ongoing passion for the research on language.

I have benefited enormously from being part of two brilliant and enthusiastic academic communities in Guangzhou and Sydney. My sincere thanks go to all my teachers in Sun Yat-sen University who have given their time and expertise to share with me their scholarly ideas on language, particularly Professors Wu Zengsheng, Lin Yuyin, Wang Bin, and Ou Hong. A huge debt of gratitude is owed to all the members of the Functional Linguistics Institute in the School of Foreign Languages, from whom I have received tremendous support. I would especially like to acknowledge the precious advice and enormous help from Professor Chang Chenguang, who kindly took over the supervision during Professor Huang’s visit to Stanford University. Moreover, the cotutelle PhD program would have been next to impossible without his committed support. Special thanks also go to Professor Dai Fan, Professor Zeng Lei, Dr Ding Jianxin, and Wang Hongyang for sharing with me their interest in multimodality and offering very helpful comments on my work. I would also like to express my gratitude to Professors He Wei, Li Guoqing, He
Hengxing, Yang Bingjun, Li Fagen, Si Xianzhu, Xiao Haozhang, Dr Yu Hui, Dr Shang Yuanyuan, Dr Wang Peng, Dr Wang Yong, Dr Deng Renhua, Dr Qi Xi, Dr Liao Haiqing, Dr Wu Kerong, Dr Yang Rufu, Dr Liao Yiqing, Dr Zhu Shan, Dr Zhang Peiwen, Dr Wang Jin, Dr Lü Dairong, Dr Long Minghui, Huang Zhiying, Xiong Xiyuan, Chen Xiaoyan, Chen Yang, Li Manliang, and Xu Yi. The importance of their timely advice and help in many aspects can never be overstated. I am very grateful to Professor Liu Yi, Wang Hongyang, and Chen Jing, who are my intellectual companions on this PhD journey. I owe a great deal to their friendship. Sincere gratitude is also due to Professor Zhang Zuo of University of International Business and Economics and Professor Qin Xiaoyi of Hainan Normal University, for their generous help and helpful advice.

Another significant aspect about the Functional Linguistics Institute for which I feel grateful forever is the opportunities it provides for us to discuss in person our research with the great figures in linguistics, including Professors Michael Halliday, Ruqaiya Hasan, J. R. Martin, Robin Fawcett, and Christian Matthiessen. Their encouraging feedback on my research proposal has given me the confidence to initiate this study.

I am equally indebted to my valued colleagues and friends in Sydney for their intellectual as well as moral support, particularly the brilliant group of researchers whose active members include Dr Chris Cléirigh, Dr Monika Bednarek, Dr Michele Zappavigna, Dr Sally Humphrey, John Knox, Zhao Sumin, Helen Caple, Naomi Knight, Ken Tann, David Caldwell, Tian Ping, Zhila Bahman, Song Chengfang, Kesuma Bakar, Chen Jing, Zhang Peiwen, Lai Liangtao, and Dominque Manghi. They have been so generous sharing with me their stimulating ideas through helpful comments and masterful presentations, which I enjoyed so much. Even the informal exchange of meanings in Royal Pub or Transient tea room has been a priceless source of motivation and information. I want to give very special thanks to Zhao Sumin, who has made painstaking efforts to scan and circulate volumes of research materials, which enabled me to commence my research on multimodality before landing in Sydney. I would like to extend my heartfelt gratitude to the scholars who have
offered critical and practical comments on my work at research seminars, postgraduate workshops, International Systemic Functional Congress, and International Free Linguistics Conference. Although it might not be possible to name all of them here, I want to thank especially Dr Sue Hood, Dr Clare Painter, Professor Theo van Leeuwen, Dr Kay O’Halloran, Professor Frances Christie, Dr Maree Stenglin, Dr David Rose, Dr Peter White, Dr Emilia Djonov, and Dr Dorothy Economou for the extra time they have taken to give enlightening advice on some chapters in this thesis. My gratitude is also due to Dr Ahmar Maboob, Dr Maree Stenglin, Jeaney Yip, Dr Michael Paton, Dr Ranjit Voola, with whom I worked as a literacy tutor and a research assistant on a collaborative EAP project (with the City University of Hong Kong) and two textbook projects for the Faculty of Economics and Business. These experiences have provided valuable opportunities for me to learn how to balance work and research.

My sincere thanks also go to the warm-hearted staff in the postgraduate offices of both institutes for their infinite generosity. They are Ms Zhong Weizhen, Ms Zhang Xueying of the School of Foreign Languages, Sun Yat-sen University, Mr Mark Try and Mr Josh Boxx of Faculty of Arts, and Ms Amanda Sayan and Mr Kenny Lowe of the International Office, The University of Sydney, without whose help I could not have solved a number of technical problems as the first cotutelle PhD candidate of the program.

I appreciate very much the financial support from different institutions at various stages of thesis production, especially the funding provided by China Scholarship Council that sponsored my research in Sydney, and the funding under The University of Sydney Postgraduate Research Support Scheme which enabled me to present papers in conferences including the International Systemic Functional Congress. Special thanks also go to Aunty Ng and her family, without whose generous help I could not have settled down in Sydney so quickly.

Finally, I want to thank the closest and dearest for their immense emotional support over the years. I am so thankful to my husband Deng Jingsong, who shares all my joy and tears in Sydney as well as in Guangzhou. Without his constant loving
support, my research would not have been so much fun. For all his unfailing support and uncomplaining tolerance even when we were oceans apart, I want to say ‘I am blessed because I am loved by you’. My deepest gratitude is due to my mother Guo Huiying, whose love is my everlasting source of courage to fight back against obstacles that seemed unsurmountable. Perhaps she will never fully understand what I am doing, but always so strongly believes in me. To her I owe a greater debt than I can possibly express. Most especially, this thesis is dedicated to my father Chen Jianzhong, who was a devoted family man and a committed EFL teacher but never had a chance to see me grow up. It is his lifelong career that I take as the focus of my study, and will cherish forever and ever.
# Table of Contents

Declaration ................................................................. ii

Abstract ................................................................. iii

摘 要 ................................................................. v

Acknowledgements ..................................................... vii

List of Figures ......................................................... xviii

List of Tables ......................................................... xix

Abbreviations ......................................................... xx

1 Introduction .......................................................... 1

   1.1 Introductory remarks ........................................... 1
   1.2 Multimodality and semiotics .................................. 1
   1.3 The interpersonal aspect of EFL pedagogic context ........... 5
   1.4 Data and methodology ........................................... 7
   1.5 Aims and research questions of the present study ............. 8
   1.6 Theoretical framework .......................................... 9
   1.7 Organization of the thesis ...................................... 11

2 Research Background: Multimodality, Multiliteracies, and EFL Education in China ........................................ 13

   2.1 Introduction ....................................................... 13
2.2 Multimodality and multiliteracies

2.2.1 Multimodality in educational context

2.2.1.1 Mediated discourse perspective

2.2.1.2 Social semiotic approach

2.2.2 Towards an understanding of multiliteracies

2.2.2.1 The concept of multiliteracies

2.2.2.2 Other approaches to multiliteracies

2.3 Multimodality research, textbook studies and curriculum standards in China

2.3.1 Development of multimodality research in China

2.3.2 Textbook studies in China

2.3.3 The interpersonal dimension in Curriculum Standards for English

2.4 Summary

3 Theoretical Foundations

3.1 Introduction

3.2 Some basic tenets of Systemic Functional Linguistics

3.2.1 Language as social semiotic

3.2.2 Stratification within and between semiotics

3.2.3 Realization and instantiation

3.2.4 Metafunctional diversification

3.3 Semantic system of APPRAISAL within Systemic Functional Linguistics

3.3.1 ATTITUDE

3.3.2 ENGAGEMENT

3.3.3 GRADUATION

3.4 Social semiotic approach to visual communication

3.4.1 Representation
4 Heteroglossic Harmony: Multimodal ENGAGEMENT Resources and Voice Interaction

4.1 Introduction...........................................................................................................86

4.2 Identifying multiple voices in multimodal EFL textbook discourse........86
   4.2.1 Previous studies on voice...........................................................................87
   4.2.2 Multi-voicing in EFL textbook discourse............................................90

4.3 Multimodal resources as ENGAGEMENT devices.................................93
   4.3.1 Labelling...................................................................................................93
   4.3.2 Dialogue balloon.......................................................................................96
      4.3.2.1 Lending support to editor voice.........................................................96
      4.3.2.2 Explaining rules of games by demonstration.................................97
      4.3.2.3 Giving directions to reader.............................................................100
   4.3.3 Jointly-constructed text.................................................................102
      4.3.3.1 Jointly-constructed drawing exercise..............................102
      4.3.3.2 Multimodal jointly-constructed herald page...............104
      4.3.3.3 Jointly-constructed verbal text with accompanying images..........107
4.3.3.4 Readership construal………………………………………109
4.3.4 Illustration………………………………………………………..112
  4.3.4.1 Image as the link between different parts of a verbal text……………………………………………………..112
  4.3.4.2 Illustration of improper behaviours………………….113
  4.3.4.3 Background and foreground illustration……………….114
4.3.5 Highlighting……………………………………………………...115
  4.3.5.1 Highlighting in visual image…………………………..116
  4.3.5.2 Highlighting in verbal text……………………………..117
4.3.6 Gradability of ENGAGEMENT values in multimodal discourse…..118
4.4 Voice interaction in multimodal EFL textbook discourse……………..…122
  4.4.1 Focalization and eye contact…………………………………..…122
  4.4.2 Reader involvement and social distance…………………………123
  4.4.3 Character engagement and point of view………………………..125
4.5 Summary………………………………………………………………….126

5 Attitudinal Accumulation: Verbiage-image Complementarity and Co-instantiation…………………………………………………………129

  5.1 Introduction…………………………………………………………………..129
  5.2 Attitudinal meanings in multimodal discourse………………………….129
    5.2.1 Social semiotic approach to verbiage-image relations………………130
      5.2.1.1 Relations between components of a multimodal text….130
      5.2.1.2 Relations between whole multimodal texts……………….132
    5.2.2 Inscribed and invoked ATTITUDE………………………………132
      5.2.2.1 Verbal inscription and invocation………………………132
      5.2.2.2 Visual inscription and invocation………………………..133
  5.3 Co-instantiation and complementarity in encoding AFFECT and
Interpersonal Meaning in Textbooks for Teaching English as a Foreign Language in China

JUDGEMENT........................................................................................................135

5.3.1 Cartoon....................................................................................................136
  5.3.1.1 Inscribed happiness of English-learning.........................136
  5.3.1.2 Gradability in affectual inscription.................................140
  5.3.1.3 Logogenetic recontextualization.................................143
  5.3.1.4 Identification or wishful thinking...............................147

5.3.2 Portrait and photograph......................................................................148
  5.3.2.1 Co-instantiation and putative reading.........................148
  5.3.2.2 Co-instantiation in JUDGEMENT invocation.............154

5.4 Co-instantiation and complementarity in encoding APPRECIATION.......158
  5.4.1 Invoked APPRECIATION of reaction.................................159
  5.4.2 Co-instantiation in encoding APPRECIATION......................160
  5.4.3 Co-instantiation in jointly-constructed text.......................166
  5.4.4 A note on degrees of commitment in visual communication.....169

5.5 The accumulation of institutionalized feelings in EFL education.......170
  5.5.1 Emotion and attitude goal in the current curriculum standards...171
  5.5.2 JUDGEMENT and APPRECIATION as institutionalized AFFECT....174

5.6 Summary..................................................................................................176

6 Contestable Reality: A Multilevel View on Modality in Visual Communication........................................................................................................179

6.1 Introduction.............................................................................................179

6.2 Modality in multimodal discourse......................................................179
  6.2.1 Linguistic modality.................................................................180
  6.2.2 Social semiotic approach to modality...................................182
  6.2.3 Modality in modes other than language...............................183

6.3 Visual modality in EFL textbook discourse.......................................186
6.3.1 Modality markers in naturalistic standard .......................... 186
   6.3.1.1 Colouration .................................................. 186
   6.3.1.2 Articulation of detail ...................................... 187
   6.3.1.3 Depth, illumination, and brightness ....................... 189

6.3.2 Coding orientations in other contexts ......................... 191
   6.3.2.1 Technological coding orientation ......................... 192
   6.3.2.2 Sensory coding orientation ............................... 193
   6.3.2.3 Abstract coding orientation ............................ 194

6.4 A multilevel view on visual modality in EFL textbook discourse . 196
   6.4.1 Sensory coding orientation in We Love Animals .............. 196
   6.4.2 Abstract-sensory coding orientation in Why Do You Like Koalas 198
   6.4.3 Naturalistic coding orientation in Wildlife Protection ........ 201

6.5 Cartooning and childist ideology .................................. 203
   6.5.1 Embodiment of childist ideology ............................ 206
   6.5.2 Relevant educational theorizing ............................. 209

6.6 A note on visual style within a macrogenre ...................... 216
   6.6.1 Teaching unit as macrogenre ................................ 217
   6.6.2 Photograph for representing actual scene .................. 218
   6.6.3 Drawing for illustrating mythology ........................ 222

6.7 Summary ........................................................................ 228

7 Conclusions and Suggestions ............................................ 231

7.1 Introduction .................................................................... 231

7.2 A retrospect of the research findings .............................. 231
   7.2.1 Recognition of multimodal ENGAGEMENT devices and their gradability ................................. 232
   7.2.2 Intermodal construal of the ‘emotion and attitude’ goal .......... 234
7.2.3 Influence of register variables on coding orientation……….235
7.3 Contributions and implications........................................237
  7.3.1 Theoretical contributions.....................................237
  7.3.2 Pedagogic implications......................................238
7.4 Suggestions for future studies........................................240
  7.4.1 Multimodality research in educational context..........240
  7.4.2 Other aspects of a metafunctionally diversified view....240
  7.4.3 Inspiration from other contexts.............................241
7.5 Envoi.............................................................................242

References.............................................................................243

Appendices.............................................................................270
List of Figures

Figure 3-1  Stratification
Figure 3-2  Language and its semiotic environment
Figure 3-3  Language’s stratified content form in relation to a stratified model of social context
Figure 3-4  Cline of instantiation
Figure 3-5  Types of structure in relation to modes of meaning
Figure 3-6  Functional diversification of language and social context
Figure 3-7  An overview of APPRAISAL resources
Figure 3-8  The ENGAGEMENT system
Figure 3-9  Interactive meaning in images
Figure 3-10  Visual focalization
Figure 4-1  Multiple voices in multimodal EFL textbook discourse
Figure 4-2  FORCE of three types of dialogue balloon in realizing [attribute] value
Figure 4-3  FOCUS of three types of jointly-constructed text in encoding [attribute] value
Figure 4-4  FORCE of two types of illustration in encoding [attribute] value
Figure 4-5  Voice interaction in multimodal EFL textbook discourse
Figure 5-1  Verbal strategies for inscribing and evoking ATTITUDE
Figure 5-2  Recontextualizing feeling in cartoon
Figure 5-3  JUDGEMENT and APPRECIATION as institutionalized AFFECT
Figure 6-1  Modality scale for colour saturation
List of Tables

Table 2-1  The world that a pedagogy of multiliteracies needs to address
Table 3-1  Levels or strata of language
Table 3-2  The gradability of attitudinal meanings
Table 3-3  The gradability of engagement values
Table 3-4  Multifunctionality for the modalities of verbiage and image
Table 4-1  Cline of instantiation – evaluation
Table 4-2  Speech functions and their corresponding visual patterns in Example (5)
Table 4-3  Cline of instantiation – from system to reading
Table 5-1  The distribution of visual styles in EFL textbooks for different levels of education
Table 5-2  Types of AFFECT in the cartoons of Example (1)
Table 5-3  ATTITUDE analysis of the verbal texts in Example (5)
Table 5-4  Inscribed and invoked ATTITUDE in the verbiage of Example (6)
Table 5-5  Inscribed and invoked ATTITUDE in the verbiage of Example (8)
Table 5-6  Inscribed and invoked ATTITUDE in the verbiage of Example (9)
Table 5-7  Emotion and attitude goals for Grades 2, 5, and 8 in *Curriculum Standards for English*
Table 6-1  The arena of the pedagogic device
Table 6-2  The narrative structure of *The Story of Atlanta*
Abbreviation

Textbook materials

PEP  People’s Education Press (人民教育出版社)

Linguistic terms

EAP  English for Academic Purposes
EFL  English as a foreign language
ESP  English for Specific Purposes
SFL  Systemic Functional Linguistics
TESOL  Teachers of English to Speakers of Other Languages
WP  world of the producer
WR  world of the reader
1.1 Introductory remarks

The link between multimodality, social semiotics, and pedagogic concerns forms the nexus of the present study. It aims to explore the ways in which linguistic and visual semiotic resources are deployed to construe interpersonal meaning in multimodal pedagogic materials for teaching English as a foreign language (henceforth EFL). To frame the research, this chapter will begin with a brief account of the research background, i.e. multimodal discourse analysis and the social semiotic approach, which is followed by a closer examination of the pedagogic environment where the current research is situated. Data and methodology will then be introduced, with research focus specified and theoretical underpinnings briefly explained. The chapter will conclude with an overview of thesis organization.

1.2 Multimodality and semiotics

Multiple representational and communicative modes are commonplace in our everyday exchange of messages. Relevant evidence can be found by looking into the evolution of human language as well as the reality of information transmission and distribution in the modern world. Human beings are believed to have had pictographic representation and communication long before the emergence of the earliest scripts such as the cuneiform script of the Sumerians in Mesopotamia around 3500 BC (Diringer 1968; Gelb 1963; Ong 1982: 83). Despite the fact that most of
the ancient scripts, including Mesopotamian cuneiforms, Egyptian hieroglyphics, Mayan scripts in Mesoamerica, and Chinese characters, have been developed independently of each other (Ong 1982: 85), each of these writing systems are considered to have pictographic origins (Coulmas 2003: 192-197). As Coulmas (2003: 209) comments, ‘the development of writing systems must be explained in terms of how visual signs are interpreted’. An understanding of the pictorial basis may shed light on the exploration of linguistic system.

Along with its manifestation in writing systems, the ‘inherent multimodality of language’ (Matthiessen 2007: 4) is also revealed in oral communication, as shown through ‘body language’ (e.g. gestures and facial expressions) and ‘paralanguage’ (e.g. vocal features) (Matthiessen 2007: 6). Linguistic multimodality has also been investigated in research on the protolanguage of children (For further discussion see Halliday 1975; Painter 1984). It may be inferred from language evolution and individual language development that human language has an intrinsic potential for multimodality. This orientation to multimodality in human communication has been further enhanced through the advances in modern digital and multimedia technology. Contemporary layout design and printing techniques allow pages to be produced as combinations of language, images, and diagrams, with a great variety of choices in typeface, font, and colouration. Moreover, the wide application of computer-enabled tools and the poliferation of web-based resources further promote the phenomenon of multimodality to an unprecedentedly pervasive level.

This increasingly multimodal reality has been one of the central topics for linguistically informed discourse analysis for the last decade (Baldry and Thibault 2006; Bateman 2008; Kress and van Leeuwen 2001, 2006; Lemke 1998a, 2000; Martin 2001, 2008a; Martin and Rose 2008; Norris and Jones 2005; O’Halloran 2005; O’Toole 1994; Scollon 2001; Scollon and Scollon 2004; Unsworth 2001, 2008). There is a growing consensus that the messages that draw on more than one semiotic resource do more than language alone can do, and ‘are assuming their central role in information dissemination in the modern world’ (Bateman 2008: 2; see also Thibault 2001: 294). As Kress (2000: 337) states, ‘it is now no longer possible to understand
language and its uses without understanding the effect of all modes of communication that are copresent in any text’.

Before the current discussion proceeds, the term ‘multimodality’ deserves brief clarification. The use of ‘multimodality’ here refers to the diverse ways in which a number of semiotic systems (i.e. linguistic, visual, audio, spatial, gestural, etc.) ‘are both co-deployed and co-contextualized in the making of a text-specific meaning’ (Thibault 2000: 312), or ‘the combination of different semiotic modes…in a communicative artefact or event’ (van Leeuwen 2005: 281). There is increasing awareness that the examination of meanings in multimodal discourse should focus on the interaction or ‘the simultaneous orchestration’ (Bateman 2008: 1) among different semiotic modes (see also Baldry and Thibault 2006; Lemke 1998a; Martin 2008a; Norris 2004). Detailed discussion and review of the major approaches to multimodality will be provided in Chapter 2.

Among different communicative contexts in which multiple modes are co-deployed to exert influence on people’s thinking and everyday life, multimodal pedagogic practices have claimed scholarly attention from educators and researchers. More and more communication in pedagogic context takes place through the interplay of various semiotic resources rather than by relying on a single semiotic mode in isolation (Jewitt 2002; Kress et al. 2001; Lemke 1998b). These developments have been discussed under the headings of ‘visual literacy’ (e.g. Dondis 1972; Messaris and Moriarty 2005; Seppänen 2006) and ‘multiliteracies’ (Cope and Kalantzis 2000; New London Group 1996; Kalantzis and Cope 2001; see Section 2.2.2 of Chapter 2 for an overview).

As compared with the growing body of work carried out in western educational settings, so far a limited amount of multimodality research has been conducted within the EFL context in China. Acknowledging the evolving multimodal teaching and learning environment in the present-day educational context in China, the current research examines the multimodal features of EFL textbook discourse, and will suggest ways in which discourse analysis may help textbook editors and users better recognize and interpret how meanings are construed through the co-deployment of
linguistic and visual meaning-making resources.

In order to systematically describe and explain the multimodal communication in the given pedagogic context, the research presented in this thesis draws on the theoretical framework of semiotics informed by linguistic theory. To be specific, it is social semiotics (Halliday 1978; Hodge and Kress 1988; Kress and van Leeuwen 2006; van Leeuwen 2005; also conceptualized as ‘systemic functional semiotics’, see Unsworth 2008: xvi) developed by systemic functional linguists that the present study is aligned with.

There has long been a tradition of scholarship of applying semiotic theory to the study of signs. The concept of ‘semiotics’ originates from the terms *semainon* and *semainomenon* used by the Stoic philosophers in ancient Greek linguistics, which mean ‘signifier’ and ‘signified’ respectively (Halliday 1985: 3; emphasis in the original). These ideas were further developed by the Swiss-French linguist Ferdinand de Saussure two thousand years later, for whom semiotics is the study of ‘the role of signs as part of social life’ (de Saussure 1916 [1983]: 15). There have been two main traditions in semiotics, i.e. structuralism including the Eastern European tradition and the Paris school, as well as a more philosophical approach developed by the American philosopher Charles Sanders Peirce that extends the reasoning and logic in natural sciences to semiotics (Lemke Net.1.; see also Stenglin 2004: 25-33). Semioticians in both traditions have analyzed signifying systems beyond language within the culture, covering photography (Barthes 1981), theatre (Bogatyrev 1938; Honzl 1940), cinema (Metz 1974), film (Mukarovsky 1978), fashion (Barthes 1985), and architecture (Broadbent 1977; Eco 1972; Jencks 1984).

In the history of semiotic studies, however, the conception of sign has tended to be viewed as ‘isolate, as a thing in itself’ (Halliday 1985: 3). In examining the artefacts that involve more than one semiotic system, we need to comprehend not only the ways in which the componential elements are pieced together, but also what the arrangements or combinations mean in a given communicative context. To facilitate this understanding, a social theory of semiotics is needed to empower analysts to grasp the meaning of semiotic choices in relation to the context in which semiotic
configurations are embedded. Moreover, this semiotic theory needs to be able to offer ‘something like a grammar’, which as Nodelman (1988: x) suggests might enable us to best understand images in picture books. The social semiotics developed by M.A.K. Halliday (1978) and his colleagues, which is strongly associated with the linguistic theory known as Systemic Functional Linguistics (henceforth SFL), provides such a robust tool for analyzing multimodal texts. As Eggins (2004: 21) states, ‘what is distinctive to systemic linguistics is that it seeks to develop both a theory about language as social process and an analytical methodology which permits the detailed and systematic description of language patterns’ (emphasis in the original). This functional-semantic orientation of SFL affords semioticians a vantage point from which a systematic discussion of contextual demands is made feasible, and a detailed linguistic and semiotic analysis is enabled (see Chapter 3 below for further explanation). Drawing upon the systemic functional semiotic approach, the current research explores the typical semiotic configurations of meaning in multimodal textbook discourse.

1.3 The interpersonal aspect of EFL pedagogic context

The current teaching and learning environment in China has become increasingly multimodal, as is acknowledged among educators (e.g. Fu 2005). A close reading of multimodal pedagogic materials that takes into account multiple semiotic systems will hopefully enable deeper understanding of how semiotic resources may ‘multiply’ (Lemke 1998a) meaning potential, which may in turn have implications for designing and utilizing pedagogic materials.

The research undertaken here attempts to locate the semiotic examination in EFL context, and this concern has partially grown out of what is presumably lacking in textbook research in China. It is argued that EFL textbooks for primary and secondary education are far less investigated than textbooks for other school subjects (Zhang 2005: 11-12). Given the common adoption of textbooks as one of the essential teaching and learning resources for primary and secondary schooling, the
present study explores different dimensions of interpersonal meaning realized by linguistic and visual meaning-making resources in EFL textbooks.

The choice of focus on interpersonal management among other possible aspects is based on one recent influential development in curriculum standards. As stipulated by Ministry of Education, curriculum standards provide guiding principles for the editing and compiling of pedagogic materials for every school subject. The eighth curriculum reform in 2001 is recognized as ‘an important milestone in our country’s curriculum development history’ (Zhong 2006: 373), one reason for which is the unprecedented highlight of the attitudinal goal for developing students’ positive emotions in the corresponding curriculum standards for all school subjects (Zhu 2006: 193). This ‘emotion and attitude’ goal is clearly articulated as one of the essential five aspects of the overall goal in EFL education (see Section 2.3.3 in the following chapter for detailed discussion). Nevertheless, the relative lack of research on how this emotion and attitude dimension is incorporated in primary and secondary textbooks has been identified as another drawback in textbook studies in China (Zhang 2005: 11). Language teaching and learning, among many other subject areas, is acknowledged as ‘far more closely related to emotion and attitude education than other school subjects’ (Cheng 2002: 32). By investigating both the verbal component and visual medium in EFL textbook discourse, the current research is meant to show how the attitudinal concerns are realized through the visual as well as linguistic semiotic choices.

In addition to the emotion and attitude dimension emphasized in the curriculum standards, two other aspects in interpersonal management call for further exploration. In the given pedagogic context, dialogic processes are advocated throughout classroom teaching, and textbook is considered to be an essential component in this process (Chen and Ye 2006; Zhong 2006). Multimodal modes of communication may further enable diversity in editor-reader alignment. Nonetheless, the way multimodal resources can be manipulated to mediate the heteroglossic space in textbook discourse remains under-researched. Still another aspect that demands attention is the fact that certain visual choices in EFL textbook discourse are
constrained by the perceived needs of the readers. The extent to which these visual
treatments are appropriate in the given pedagogic context is yet to be fully examined.
Knowledge structure and the nature of pedagogic discourse (Bernstein 1990, 2000)
may need to be considered when assessing visual design in pedagogic materials.

The aforementioned three aspects of interpersonal meaning (i.e. editor-reader
alignment, attitudinal dimension, and visual design in relation to contextual factors)
constitute the major pedagogic concerns of the current research. Further explanation
of the interpersonal management in the given EFL education context is provided in
Chapter 2. It is hoped that by seeing more clearly how linguistic and visual semiotic
choices work in tandem to construe the dialogic setting, to hint at the intended
evaluative stance, and to respond to the discerned readers’ needs, we may arrive at a
better understanding of the functions fulfilled by linguistic and visual semiotic
systems in pedagogic context.

1.4 Data and methodology

The data of the present study comprises eighteen EFL textbooks for primary
education, junior and senior secondary education, edited and published by People’s
Education Press (henceforth PEP) between 2002 and 2006, which include an entire
series of the latest available editions at the time when the study commenced.
Altogether a hundred and eighteen teaching units are involved (i.e. forty-seven in the
primary textbooks, forty-six in the junior secondary ones and twenty-five in the senior
secondary textbooks). All teaching units contain visual images as well as verbal
texts, ranging from those multimodal texts with a relatively small amount of verbiage
for primary schooling through to those with a greater proportion of verbiage for
secondary education. Some statistical summaries show that one thousand three
hundred and ninety-eight visual images of different styles (see Table 5-1 in Chapter 5)
are employed in the EFL textbooks under discussion. The large number of visual
images involved in the textbooks indicates a pressing need to grasp the meanings
encoded in the multimodal texts for pedagogic context. EFL textbook discourse can
be viewed as a typical configuration of semiotic choices from both linguistic and visual semiotic systems. However, EFL textbooks are yet to be fully examined at the micro, discursive level (Zhang 2005: 11-12), and far less understanding has been achieved in terms of the semiotic description of their multimodal nature.

Working towards an in-depth discursive account of EFL pedagogic materials, the current research design is descriptive and qualitative in nature, though some statistical treatments will be employed to describe the status quo of the data as faithfully as possible. Owing to the limited amount of research on multimodal EFL textbooks in China, this study is exploratory and explanatory in orientation. The major theoretical rationale informing the current research is SFL, and a systemic functional semiotic view is adopted when examining multimodal excerpts from EFL textbooks. The analysis will cover both interactive and personal dimensions of interpersonal meaning, encompassing three sub-studies. The overall objective and specific research questions are provided in the ensuing section.

1.5 Aims and research questions of the present study

The ultimate goal of the current research is to develop a comprehensive theoretical account to model all aspects of the communicative function realized in the given EFL context. The focus of this thesis is on one dimension, i.e. interpersonal meaning, in EFL textbook discourse. The research findings will give rise to insights into both the interactive aspect like editor-reader alignment and the personal dimension such as the construal of evaluation. The research questions to be addressed in this thesis are as follows:

1. How are linguistic and visual meaning-making resources deployed to mediate the heteroglossic space in the multi-voiced EFL textbook discourse? In what way the degree of alignment between voices may be graded by utilizing multimodal resources?

2. How is the evaluative stance in multimodal EFL textbook discourse construed through linguistic and visual semiotic systems? What
kinds of verbiage-image relations are there in the intermodal construal of attitudinal meanings?

3. In what way are multimodal messages defined as real in the EFL textbooks targeting readers of different educational levels? How is it conditioned by as well as construing the context in which it is embedded? Is the visual arrangement in line with the nature of pedagogic discourse, and why?

As is implicated in the research questions, the present study is intended to shed light on both theoretical and pedagogic issues. A close linguistic and semiotic analysis of EFL textbook discourse may facilitate the understanding of as well as provide solid evidence for some attention-engaging concerns like ‘dialogic process’ and ‘emotion and attitude’ goal in the given pedagogic context. An interrogation of the theoretical assumptions on which the current research is based will enrich the linguistic and semiotic studies of interpersonal meaning in multimodal discourse that involves more than one semiotic system.

1.6 Theoretical framework

As was indicated above, the theoretical framework underpinning the present study is SFL (Halliday 1994; Halliday and Matthiessen 1999, 2004; Martin 1992), including the APPRAISAL system (Martin 1997, 2000a; Martin and White 2005) for modelling interpersonal semantics. Systemic functional semiotic approach (Hodge and Kress 1988; Kress and van Leeuwen 2006; van Leeuwen 2005) is drawn upon to analyze visual meaning-making resources. Two fundamental features of SFL, among many others, prove itself an efficient tool in exploring semiotic choices in EFL textbook discourse. In the first instance, it considers language as a social semiotic and defines ‘a culture as a set of semiotic systems, as sets of systems of meaning, all of which interrelate’ (Halliday 1985: 4). In other words, language together with other systems of meaning (e.g. visual, audio, gestural, etc.) constitute the human culture. Moreover, in SFL language is theorized as conditioned by as well as construing social context.
Meaning is construed in the immediate context of situation and the wider context of culture (Huang 2003; see also Huang and Ghadessy 2006). Recent advances in systemic functional semiotic theory recognize that the same meanings may often be encoded in different semiotic modes and common semiotic principles can operate in and across different modes (Kress and van Leeuwen 2001: 1-2). The functional-semantic orientation of SFL enables us to explore how semiotic systems within EFL textbooks both reflect and construct the language teaching and learning environment in which the textbook discourse is embedded.

In SFL, interpersonal meaning is one of the three metafunctions that are simultaneously construed whenever language is used. Situated within the broader model of SFL, the APPRAISAL system offers a functional model of interpersonal meaning at the level of discourse semantics, which may address the analysis demand arising from some current pedagogic concerns, including editor-reader alignment and construal of evaluative stance. The sub-systems and options within APPRAISAL are semantic categories, and as such the realizations of these categories may transcend monomodal verbal communication to include multimodal meaning-making resources. Together with systemic functional semiotic approach to visual images, the APPRAISAL network has the potential to bring together verbal as well as visual choices that might sometimes be left out when considering values and voices in a text.

In addition to the application of linguistic and semiotic theory, the current research also aims to make a contribution to the functional theory it draws upon. The investigation into the resources that realize ENGAGEMENT, GRADUATION, and ATTITUDE meanings in a multimodal context will lead to an expansion of the theoretical network of APPRAISAL. In addition, the degree of ‘commitment’ (Martin 2008b) and the ways in which choices from different systems may be ‘coupled’ (Martin 2008b) with one another in multimodal texts will also be considered. A detailed explanation of the theoretical foundations will be provided in Chapter 3, and further elaboration will be presented as the analyses unfold in Chapters 4 to 6.
1.7 Organization of the thesis

This thesis is divided into seven chapters. This introductory chapter has outlined the research background and explained the general research design. Chapter 2 further clarifies the notions of multimodality and multiliteracies, reviewing the existing body of work in the research area. Relevant textbook studies and major themes in *Curriculum Standards for English* in the present-day China are outlined, with research gaps identified and focus of the present study framed. Chapter 3 introduces the theoretical rationale of SFL that will be drawn upon in the ensuing analyses, particularly the semantic system of APPRAISAL and the systemic functional semiotic approach to visual analysis. The latter part of the chapter provides the theoretical justifications for the analyses to be conducted, and serves as a transition to the three sub-studies in the subsequent chapters.

Detailed analyses and discussions of the different aspects of interpersonal meaning in EFL textbook discourse begin in Chapter 4 and continue through Chapter 6. Chapter 4 is devoted to the interactive dimension of interpersonal management, examining how linguistic and visual semiotic resources can be deployed to align the ‘ideal’ reader with the propositions and values advanced in a text. Attempts will be made to explore and extend the semantic systems of ENGAGEMENT and GRADUATION by uncovering a range of ENGAGEMENT devices in multimodal communicative context, with a detailed description of the ways in which these multimodal resources mediate the heteroglossic space in EFL textbook discourse. The interplay between editor voice, reader voice, and character voice will also be considered with reference to contact, social distance, and point of view. Chapter 5 focuses on the intermodal construal of evaluative stance, investigating the ways in which verbiage and image either complement or co-articulate with each other in establishing the attitudinal orientation for different levels of education. A diversity of visual styles, ranging from cartoon to portrait and photograph, are investigated along with the corresponding verbal texts. With regard to the ‘emotion and attitude’ goal in curriculum standards, an ontogenetic view will be developed to discuss the attitudinal
shift that happens as students advance through the school years from primary to secondary education. Chapter 6 looks into the concept of modality in multimodal discourse, which is concerned with how true a given representation is supposed to be taken. Multimodal texts from primary, junior and senior secondary textbooks will be compared, in order to provide an insight into the interdependent relationship between coding orientation and social solidarity. In relation to the educational theory on pedagogic discourse and nature of knowledge, a critical view on visual style in pedagogic context will be offered. The phenomenon of various visual styles adopted within a macrogenre leads to a discussion of the influence of field on visual treatments.

Chapter 7 summarizes the major findings of the research and its contributions to both linguistic studies and pedagogic agenda. Suggestions for future research will also be indicated in the final section of the concluding chapter.

NOTES

1 The present study follows the view in social semiotics that the terms ‘discourse’ and ‘text’ represent two complementary perspectives on the same phenomenon. Specifically, ‘discourse’ is used to refer to ‘the social process in which texts are embedded’, while ‘text’ is adopted when mentioning ‘the concrete material object produced in discourse’ (Hodge and Kress 1988: 6).

2 This thesis follows SFL conventions (e.g. Martin 2000a; Kress and van Leeuwen 2006) in using small capitals to indicate systems (e.g. ENGAGEMENT, APPRAISAL), square brackets for the features of a particular system (e.g. [proclaim], [attribute]), and initial letters capitalized when talking about functional labels (e.g. Actor, Goal).
Chapter 2

Research Background:
Multimodality, Multiliteracies, and EFL Education in China

2.1 Introduction
In this chapter, I shall focus on both the linguistic and pedagogic landscape in which the present study is positioned, aiming to offer a better understanding of the current research topic and research background. First I shall introduce the concepts of multimodality and multiliteracies. Some of the major literature that represents different approaches to the research area of multimodality will be reviewed in detail, within which research gaps will be identified and from which inspiration will be drawn.

Following the brief introduction of other approaches to multiliteracies research, a general account of the present-day scenario of the multimodality research and textbook studies in China mainland will be provided, whose limitations partially explain the necessity of the present study. The curriculum standards that exert influence on EFL textbook-editing will then be discussed, with a particular focus on the interpersonal dimension.

2.2 Multimodality and multiliteracies
With the progress in modern communications technology, people’s interaction and exchange inevitably draw upon channels and means that involve more than a single semiotic mode, and hence the notion of ‘multimodality’ emerges to cover the
correlated changes in discourse studies. One of the theoretical advances conceptualizes multimodality as ‘the combination of different semiotic modes…in a communicative artefact or event’ (van Leeuwen 2005: 281). The educational setting, which is the central concern of the current research, is one of the major contexts in which multimodal communication occurs. The influence of multimodality on pedagogic practices has captured scholarly attention. Along with those studies dealing with cultural and linguistic diversity, the approaches concerning the multiplicity of communications channels and media constitute what is termed as the research on ‘multiliteracies’ (New London Group 1996; Cope and Kalantzis 2000).

2.2.1 Multimodality in educational context

In the exploration of multimodal discourse in relation to modern technology of printing, computing and multimedia, two most influential trends are recognized. One of them takes a ‘mediated’ discourse perspective, while the other draws on linguistic and social semiotic theory. The present section aims at outlining principles of the two approaches, with focus placed on their respective studies in educational context. In spite of the fact that both perspectives have put the investigation of multimodal resources on the agenda, the one that advocates ‘a nexus of practice’ seems to emphasize ‘mediated’ social actions, whereas the tradition that is more relevant to the current research grounds its examination of multimodal texts in the analytical underpinnings that are informed by linguistic theory.

2.2.1.1 Mediated discourse perspective

Those researchers who advocate ‘mediated discourse analysis’ (Norris and Jones 2005; Scollon 2001) view discourse analysis as social action, or as Scollon and Scollon (2004: 7) put it, ‘use discourse analysis to engage in social action’. In other words, the concern of mediated discourse analysis is not discourse per se but social action. Their approach is often termed as ‘nexus analysis’, which is ‘a form of ethnography that takes social action as the theoretical center of study’ (Scollon and Scollon 2004:
In mediated discourse approach, the smallest, basic unit of analysis is the so-called ‘mediated action’ (Scollon 1998), which was first put forward by Wertsch (1991, 1998) and interpreted with the emphasis that ‘any action is inherently social’ and ‘carried out via material and symbolic mediational means (cultural or psychological tools)’ (Scollon and Scollon 2004: 12). Mediated action is the focus of mediated discourse analysis, and ‘must always remain problematized’ (Scollon 1998: 11). There is an inherent irreducible tension between its two elements, i.e. agent and mediational means (Norris and Jones 2005: 17). Mediational means refer to the semiotic means through which mediated actions are conducted. As Scollon (2001: 14) explains, mediational means are ‘neither external objects nor internal psychological constructs alone but rather are a dialectical relationship between objective materiality and psychonological or intramental process’. In addition to the focus on mediated action, mediated discourse analysis also stresses the importance of social actors as they are acting, because ‘these are the moments in social life when the Discourses which we are interested in are instantiated in the social world as social action’ (Scollon 2001: 3, capitalization in the original).

Mediational means, social actors and the sociocultural environment intersect at the real time moment of mediated action. The ‘concrete, real-time’ mediated actions with a historical accumulation within the habitus of social actors are recognizable to and understood by other social actors to be the social actions of the same type, and thus the concept of ‘practice’ is used for referring to this kind of mediated action. In addition to the concepts of mediated action, mediational means, and practice, other important technical terms within mediated discourse analysis also include ‘site of engagement’ and ‘agency’. All mediated actions are regarded as taking place within a site of engagement. According to Scollon (2001: 3-4), a site of engagement is defined as ‘the real-time window that is opened through an intersection of social practices and mediational means (cultural tools) that make that action the focal point of attention of the relevant participants’. Norris (2002) further develops this concept by introducing the notion of different levels of action and the continuum from focused
action to less focused and to unfocused actions of participants. As Norris and Jones (2005: 140) state, one of the difficulties of operationalizing the concept of sites of engagement is locating them in time, and the discussion of sites of engagement should also cope with how the researchers contribute to the construction of such sites through their own actions, as well as the convergence of multiple trajectories of actions and practices from both the researchers and what is being researched in the sites of engagement. The concept of ‘agency’ within mediated discourse analysis is viewed as integrated in tension with the social actor’s habitus, the mediational means available and the social practices involved in the mediated action construction. Agency is always ‘distributed’, and there is a constant negotiation between individuals and the social world. However, the questions of how exactly this negotiation takes place and what is the role of agency play in the production of social actor and social world need further studies (Norris and Jones 2005; see also Scollon 2001). Drawing on Burke’s (1969) view of agency as a matter of perspective, mediated discourse analysts examine agency in relation to the issue of power and domination, the tension between its construction and interpretation, and the researchers’ own observation and their horizon of awareness (Norris and Jones 2005: 170).

In mediated discourse analysis, larger activities are viewed as ‘nexus of practice’ (Scollon 2001), and ‘nexus analysis’ (Scollon and Scollon 2004) is the methodological principles of mediated discourse analysis. According to Scollon and Scollon (2004: 9-10), nexus analysis consists of three main steps. Firstly, researchers should identify themselves as part of the nexus of practice they aim to study, and this task is named as ‘engaging the nexus of practice’. The second step is ‘navigating the nexus of practice’, where the researchers begin the data collection and analysis. It is the main body of a nexus analysis and the scope of study largely depends on the ‘circumferencing’, which is ‘the analytical act of opening up the angle of observation’. The third activity is termed as ‘changing the nexus of practice’, which aims at producing social changes. From the start, mediated discourse analysis takes into consideration the subjectivity of researchers and participants, and hence a
constant reflective process is required, which is termed as establishing the ‘zone of identification’ (Scollon and Scollon 2004). The second step of nexus analysis is concerned with data processing. The data in mediated discourse analysis are collected in different modes of communication and from various points of view, which summarized as the distinctive features of ‘multimodal’ and ‘multiperspective’ by mediated discourse analysts (Norris and Jones 2005: 202). Furthermore, there is a series of analytical processes guiding the data collection, including the analysis of persons, discourse, other meditational means, trajectories, timescales, and motives. The final step of nexus analysis echoes the goal advocated by mediated discourse analysts to bring about social changes. To be precise, what is to be done at this stage is to investigate how the nexus of practice ‘already has changed’ during the research, owing to the fact that the process of change actually begins in the first step when the researchers enter the nexus of practice (Norris and Jones 2005: 203).

Mediated discourse analysts set the goal of understanding and solving social problems by bringing together principles from diverse disciplines such as ethnomethodology, conversation analysis, linguistic anthropology, psychology, and sociolinguistics. Researchers in this field have covered topics including AIDS prevention (Jones 1999, 2002), immigration (Johnson 2002), national identity (de Saint Georges and Norris 1999), childhood literacy (Boswood 2000; Scollon et al. 1997), news discourse (Scollon 1998), and gender (Norris 2005). Within educational setting, mediated discourse analysis concerns itself with the technologies of literacy ranging from traditional print literacy to the use of electronic communication and the internet (Scollon 2001: 10). For instance, Scollon and Scollon (2004) analyze the social actions of teachers and students involved in two university classes. One of them is a traditional, face-to-face classroom centring around the reading and the production of essays; whereas the other is the ‘technologically mediated class’ that is mediated by technological means of email and audio conferencing. As Scollon and Scollon (2004: 15-16) argue, the analysis of the two distinct ‘educational nexus’ illustrates the way in which technology as a meditational means ‘redistributes the interaction order, brings some new discourses
Research Background: Multimodality, Multiliteracies, and EFL Education in China

into play while setting other ones into the background, and is differentially established in the historical bodies of the participants’. Based on the findings they contend that the technologically mediated class promotes the development of new forms of communication while restraining traditional forms of university classroom interactions. As is suggested by Scollon and Scollon (2004: 16), access to university instruction is thus ‘redistributed in ways that come to serve different social goals, purposes, and groups’.

When it comes to the methodological principles, mediated discourse analysis seems to start with an ethnographic orientation and later allows for interdisciplinarity. As Scollon and Scollon (2004: 9) articulate, ‘a nexus analysis is an ethnographic methodological strategy’. They further interpret ethnography as ‘an extended study of action(s) undertaken by people in the course of living their lives’ and stress the importance for an analyst ‘to be identified within the nexus of practice under study’ (Scollon and Scollon 2004: 13). Mediated discourse analysts treat nexus analysis as ‘a set of heuristic tools’ with which analysts can bring in ‘any data collection and analytical tools’ as long as they are considered useful in dealing with social actions (Norris and Jones 2005: 201). It may be justified to say that, as far as methodology goes, one of the major differences between mediated discourse analysis and the approach informed by linguistic theory (see Section 2.2.1.2 below) lies in the fact that the former is not be grounded within linguistic or semiotic framework. Moreover, what is of relevance to mediated discourse analysts tends to be the ‘action’ rather than the linguistic or semiotic features involved in it (see Scollon 2001: 17).

To sum up, mediated discourse analysis itself is deemed as a ‘nexus of practice’ that links many different theoretical and methodological perspectives from various disciplines in social science (Norris and Jones 2005: 204). Setting out with an ethnographic position, mediated discourse analysts call for disciplinary diversity in studying social actions and believe that broader socio-political-cultural analysis can be grasped through micro-analysis of unfolding moments of social actions (Scollon and Scollon 2004: 8). Nevertheless, as is pointed out by Norris and Jones (2005: 9), mediated discourse analysis approaches the question of how discourse is a matter of
social action through the action rather than through the discourse. In other words, it expresses the concern with social action before discourse, and examines the role of discourse in action only after figuring out what the action is.

2.2.1.2 Social semiotic approach

Another foundation for the account of multimodal meaning-making resources derives from Systemic Functional Linguistics and social semiotic theory (Halliday 1973, 1978, 1994; Halliday and Matthiessen 2004; Hodge and Kress 1988; van Leeuwen 2005). The significance of research on multimodality has long been recognized by linguists and social semioticians. As is pointed out by Hodge and Kress (1988: vii-viii), meaning resides so strongly and pervasively in semiotic systems other than language that a concentration solely on verbiage is far from being sufficient. From the perspective of social semiotics, the order of the world ‘is expressed semiotically through choices from a variety of sign systems including language, visual images, music, gesture, and three-dimensional objects’ (O’Halloran 2005: 6). Beginning with the discussion of the ‘grammar’ of visual images (Kress and van Leeuwen 2001, 2002, 2006; O’Toole 1990, 1994), linguists and social semioticians’ exploration of the semiotic choices in multimodal texts has been extended to cover other semiotic systems such as sound and music (van Leeuwen 1999), action (Martinec 1998, 2000), spatial design (Martin and Stenglin 2007; Ravelli 2006; Stenglin 2004), electronic media and film (Baldry 2004; Baldry and Thibault 2006; Thibault 2000; Zammit and Callow 1999), and architecture (O’Toole 2004). Their analytical framework, especially the one for systematically modeling visual images, will be accounted for in detail in the ensuing chapter. In the current review I shall focus on the foundational and recent research on multimodality conducted by systemicists and social semioticians, with particular focus on those studies concerning the educational context.

When reflecting on the ‘action research’ that has been carried out by the ‘Sydney School’ ever since the early 1980s, Martin (2000b) points out that the new
millennium witnesses the fifth phase of the research in the context of education and workplace, which is characterized as multiliteracies. The SFL contributions to the study of multimodal teaching and learning resources cover various semiotic modes of communication ranging from print and electronic media to classroom interaction, and from English literacy to other school subjects. I shall review some of the recent advances on this frontier in the following three aspects: scientific printed texts, classroom practice, and electronic media.

(i) **Scientific printed texts** Systemicists and social semioticians have carried out substantial studies on multimodal printed texts, among which those focusing on scientific texts in both educational and professional research contexts constitute a large proportion. The topics that have been covered in this research area encompass modern scientific discourse, ecology discourse, mathematical discourse, biology discourse, and the relations between different genres in science. Lemke (1998a) studies the meanings that arise from the interaction between different semiotic systems in modern scientific texts. For Lemke, the metafunctional diversifications of ‘presentational’, ‘orientational’ and ‘organizational’ meanings provide the possibility of analyzing co-existing semiotic systems (language, figures, tables, visual images, etc.) in scientific texts. In his terms, these semiotic systems are ‘incommensurable’ in the sense that the meaning constructed by one semiotic modality (e.g. image) cannot be constructed in exactly the same way by another semiotic modality (e.g. verbal text). He further points out that the interaction between different semiotic resources gives rise to a ‘multiplying’ effect (Lemke 1998a: 92). Veel (1998) examines the linguistic features in texts on ecological science in comparison with those on ‘traditional science’. His findings show that in addition to the lexico-grammatical patterns already found in traditional science, other linguistic features in eco-scientific texts also include register metaphor (Martin 1997), humans as grammatical agents, and the tension between theoretical laws and actual physical world that creates the ‘humanities-like rhetoric’ effect. Drawing on Kress and van Leeuwen’s (1996) visual analytical framework, Veel also demonstrates that visual displays in eco-scientific texts prove to be a powerful resource for construing the
relationship between human beings and environment.

Mathematical discourse, according to O’Halloran (1999a, 2005), is necessarily multisemiotic, in the sense that the three semiotic resources involved in mathematical discourse (i.e. mathematical symbolism, visual display, and language) are functional sign systems that differentially construct reality. O’Halloran proposes a systemic framework to describe the meaning potential (i.e. experiential, logical, interpersonal, and textual meanings) of mathematical symbolism and the role of visualization in mathematics. When examining the intersemiosis between various semiotic modes, O’Halloran introduces the notion of ‘semiotic metaphor’ to capture the phenomenon where an intersemiotic transition leads to a metaphorical expansion of meaning. Guo (2004) works on the multisemiotic nature of introductory biology textbooks. Drawing on systemic functional approach to multimodal scientific discourse (O’Toole 1994; Lemke 1998a; O’Halloran 1999a) as well as sociological studies of biology texts, Guo provides a social semiotic framework for analyzing schematic drawings and mathematical graphs in biology, demonstrating how the various semiotic resources interact with each other to make meanings. He argues for paying due attention to the visual as well as linguistic meaning-making resources in EAP teaching, stressing again the important role of ‘visual/textual interactivity (VTI)’ (Johns 1998: 186) in the field of science education.

On the level of genre, Martin and Rose’s (2008: 139-227) research offers a typological view on the genre relations within scientific discourse. The different genres identified in the field of science include report, explanation, procedure, and procedural recount. Their discussion is extended to cover visual images in science (Martin and Rose 2008: 167-177). The ideational meaning construed in technical images is systematically investigated along the axes of ‘phonomenon focus’, ‘category’, and ‘representation’. In addition, options for the textual organization of images as well as logical relations between visual images and verbal texts are explored. When relating this social semiotic exploration to the socioeconomic structure in modern industrial society, Martin and Rose (2008: 224-226) point out that the written scientific discourse becomes more remote from the construal of everyday
experience in spoken discourse along the industrial ladder and education sequence, and the unequal acquisition of the science genres may exert great influence on the social structure as well as occupational options.

(ii) Classroom practices  The multimodal social semiotic approach to classroom practice puts onto its agenda the exploration of teaching and learning resources beyond printed texts. Science classrooms have gained much scholarly attention among other subjects of classroom teaching. Through the close observation on a student’s multi-literacy practices in an advance chemistry class and an advanced physics class that involve videotapes, student notes, teacher handouts, overhead transparencies and textbook selections, Lemke (2000) outlines the multiliteracy competence that students need to develop to integrate and coordinate the specialized verbal, visual, and mathematical literacies. In doing so, students are able to reinterpret and recontextualize information in one channel in relation to that in other channels, thus meeting the maximal literacy demands of the post-compulsory scientific curriculum. Besides classroom practice, the multimedia literacy demands of scientific genres, as Lemke points out, also exert influence on curriculum design, pedagogy, assessment and research.

In the sphere of mathematics education, O’Halloran (2000) examines the multisemiotic nature of mathematics on classroom discourse from an SFL perspective, with the purpose of understanding the difficulties inherent in mathematics teaching and learning. It is shown that the three semiotic resources involved in mathematical discourse, i.e. mathematical symbolism, visual display and language, have their own unique lexicogrammatical systems for encoding meanings, and these three types of resources interact with each other to shape the construction of the pedagogical discourse in classroom practices. Particularly in the case of mathematical symbolism, meaning is encoded ‘unambiguously in the most economical manner possible’ through the use of specific grammatical strategies such as multiple levels of rankshifted configurations of ‘mathematical Operative processes and participants’. The reconfigurations of symbolic Operative processes and participants in oral discourse in mathematics classroom are also examined. O’Halloran’s research also
indicates that the dense texture of mathematics pedagogical discourse derives from the inclusion of the symbolic constructions in the linguistic metadiscourse, and what she calls ‘semiotic metaphor’ is employed to explain how shifts in meaning of functional element occur and how new entities are introduced with movements between semiotic codes.

Based on interdisciplinary observation and concrete analysis, Kress et al.’s (2001) research in science classrooms challenges the traditional view on language’s central or dominant role in teaching and learning. It is argued that other modes of communication such as image, gesture, spatial and bodily codes also contribute to the multimodal ways of meaning-making and knowledge construction. Students’ action in the science classroom is regarded as the active engagement with all the multimodal modes of communication, and all participants in communication are viewed as active transformers of the meaning-making resources around them. Overall, the multimodal approach to the science classroom practice offers a way of reconsidering the role of language in the changing reality of contemporary literacy due to multimedia and electronic technology, thus drawing attention to the new demands on the development of students’ multiliteracy competence. Also placing the emphasis on the ‘multimodal communicative competence’ that teachers should help learners to develop, Royce (2002) investigates the ‘intersemiotic complementarity’ (Royce 1998) between visual images and written language in the TESOL classroom. Through the analysis of a multimodal environmental text at high school level with particular focus on ideational meaning, Royce demonstrates that along with the meaning-making features specific to visual and verbal modes respectively, the two semiotic systems also complement each other to realize ‘an intersemiotically coherent multimodal text’ when they co-occur in a text. Besides, the intersemiotic resources for realizing this complementarity can serve the pedagogical purposes in the activities of TESOL classroom such as reading, writing, speaking, listening, and vocabulary learning.

As for the research on multimodal classroom practices in different school subjects, Unsworth (2001) suggests developing and teaching multiliteracies in early school years as well as in upper primary and junior secondary schooling. He
Research Background: Multimodality, Multiliteracies, and EFL Education in China

outlines the distinctive textual forms and literacy practices of different school subject areas including science and humanities, and explores the verbal and visual meaning-making resources in the literature for children. In terms of classroom practicalities, Unsworth draws on sample lesson materials to show how to address the three facets of critical literacy (i.e. recognition literacy, reproduction literacy, and reflection literacy) of young children, and explores the complementary use of conventional and computer-based texts in primary and secondary schooling. Unsworth further shows that the knowledge dimension, pedagogic dimension and multiliteracies dimension of classroom learning contexts are significant in dealing with the practicalities of implementing multiliteracies in learning and teaching, which provides implications for literacy educators in the design and implementation of literate activities.

(iii) Electronic media With the advances in computer and multimedia technologies, new kinds of literacy practices, i.e. computer-based or screen-based literacies, have emerged and aroused scholarly interest. As Unsworth (2001: 12) puts it, multimodality is not an exclusive feature of electronic texts, but the range of modalities, the extent of their use, and the nature and quality of their articulation, have significantly increased in electronic formats. The electronic texts under scholarly attention include those used in EAP/ESP classroom, those for English literacy purposes, and those for higher education in science. Baldry (2000) argues for an awareness of the multimodal organization of scientific texts in EAP/ESP classroom by developing a historical and comparative view on both static (i.e. printed media) and dynamic (i.e. film and computer media) multimodal texts. To be specific, he describes the changes that happened in the organization of the printed scientific page in biological science and economics over a hundred years of development, and explores the evolution of dynamic multimodal texts such as lectures and film media in medical science. In addition, he points out the necessity of specialized software tools (e.g. online multimodal corpus) in the development of multimodal syllabuses for EAP/ESP classroom.

Zammit and Callow (1999) examine the reading of multimedia CD-ROMs
through the analysis of two screens from factual programmes. They describe how users may be influenced by the programme design and need to shift their reading positions when moving between different texts on the screen. Through the comparison between the ways in which the two CD-ROM programmes construct information, Zammit and Callow argue for explicit, critical reading skills for multimedia texts, which may enhance students’ interaction with CD-ROM programmes. Also placing the focus on CD-ROMs in educational setting, Jewitt (2002) investigates the transformation of a novel to CD-ROM used in school English, with emphasis on the multimodal reshaping of the entity ‘character’ in the shift of representational modes. In doing so, she probes into the way in which multimodal resources in the CD-ROM, especially the visual ones, provide students with different possibilities for engaging themselves with specific characters in the novel. It is found that character is presented as a fluid entity rather than a stable one, thus requiring students to read in a wider social-historical context beyond the text. In terms of the implications for literacy, Jewitt suggests that the process of learning within school English should be viewed as more than a linguistic accomplishment.

Addressing the issue of ‘computer-age’ children’s multimedia digital literacy that involves CD-ROM, internet and electronic game narratives, Unsworth (2005) offers practising teachers and students of teacher education ways of understanding how digital technology may enhance children’s literary experience as well as practical suggestions on planning programs by using e-literature in classroom. Specifically, Unsworth puts forward the ‘interpretive’, ‘organizational’ and ‘pedagogic’ frameworks that enable teachers to effectively manage classroom literacy programs with digital resources. The interpretive framework describes the integrative role of language and images in computer-based literary texts as well as conventional printed books, while the organizational framework introduces the electronically ‘augmented’, ‘re-contextualized’ and ‘originated’ articulation in e-literary texts. Along with the explication of different facets of e-literature for children, Unsworth also provides examples of e-literary classroom programs targeting children at different levels of primary schooling.
In defining the goals of advanced literacy in science, Lemke (2002) explores a third domain of scientific literacy along with the traditional scientific printed publications and scientific curriculum, i.e. the internet-based multimedia genres that professional scientists design to communicate with one another and also with the general public. Through the detailed analysis of two different parts of the National Aeronautics and Space Administration (NASA) website, Lemke finds out that the multimedia literacy demands of the internet-based genres are both comparable to as well as beyond those of scientific printed publications. The emergence of hypertext literacy, i.e. the features of the extended literacy of computer-based multimedia, raises the question of how to navigate through hypertextual web networks as well as how to make meaning across sequentially and locally linked webpages.

To summarize, the approach informed by social semiotics (Halliday 1978; Hodge and Kress 1988; van Leeuwen 2005) to the literacy practices interprets ‘literacy’ in a broad sense as involving meaning-making or semiosis in general (Lemke 2002: 22). Taking the principles from Systemic Functional semiotics tradition as the theoretical underpinning of the present study, I limit the scope of the present study by focusing on the verbal and visual meaning-making resources in the EFL textbooks currently used in China, whose multimodal nature is far less understood and investigated when compared to their western counterparts. Following the discussion of the concept of multiliteracies in the following section, the remainder of this chapter will be devoted to the general scenario of multimodality research, textbook studies and the interpersonal dimension of curriculum standards in China.

2.2.2 Towards an understanding of multiliteracies

It has been widely recognized that multiple modes of representation and communication as well as diversified cultural and linguistic features are highlighted in modern discourse, which have been constantly changing in contemporary workplaces, public spaces and personal lives (New London Group 1996; Cope and Kalantzis 2000; Kalantzis and Cope 2001). The fundamental premises of literacy pedagogy need to be reconsidered so as to provide students with the skills and knowledge they need in
achieving their aspirations in this rapidly changing world. Since the mid-1990s when the concept of multiliteracies was put forward, it has attracted scholarly attention from various perspectives. Substantial studies have been carried out from the perspective of cultural and linguistic diversity as well as in the dimension of multimediated literacy. In the present section, I shall attempt to spell out the relevant technical terms in multiliteracies pedagogy, and briefly outline some major trends in this research area.

2.2.2.1 The concept of multiliteracies

One way of looking at pedagogy is to view it as the teaching and learning relationship that engenders the potential for creating the learning conditions that lead to full and equitable social participation (Cope and Kalantzis 2000: 9). Literacy pedagogy is thus expected to fulfill a socially oriented mission. As the New London Group (1996) observe, in contemporary society dramatic changes are taking place in people’s working lives, public lives as citizens, and private lives as community members. To be specific, the ‘productive diversity’ in the ‘new capitalism’ (Gee et al. 1996) calls for a new literacy pedagogy that accounts for the cross-cultural communication and negotiation across linguistic and cultural differences, so as to prepare students for the new forms of working life. As for citizenship, due to this local diversity and global connectedness a ‘civic pluralism’ has emerged, and hence the need to negotiate regional or ethnic dialects, variations in register, hybrid cross-cultural discourses and multi-semiotic meanings has become one of the important skills that students need to master. Just as worklife and civic spaces have changed, so has the personal life of individuals. With the increasing invasion of private spaces by the mass media and information networks, private lives are rendered more public because everything private can be a potential topic for media discussion. That leads to the widespread ‘conversationalization’ of public language (Fairclough 2000), and people’s identities have multiple layers with complex relations in between. Besides, multimedia and hypermedia technologies enable members of a given culture/subculture to have their voices in this multiplicity of lifeworlds.
In response to the increasingly divergent lifeworlds in the aforementioned three realms, the notion of multiliteracies has been proposed to take cultural differences into account and view language and other modes of meaning as dynamic, constantly remade in the changing contexts (New London Group 1996). Viewing curriculum as ‘a design for social futures’, the New London Group outline the changing realities (see Table 2-1 below) that a pedagogy of multiliteracies needs to take into account.

Table 2-1 The world that a pedagogy of multiliteracies needs to address
(New London Group 1996)

<table>
<thead>
<tr>
<th>Working lives</th>
<th>Public lives</th>
<th>Personal lives</th>
</tr>
</thead>
<tbody>
<tr>
<td>Changing realities</td>
<td>Decline of the civic</td>
<td>Invasion of private spaces</td>
</tr>
<tr>
<td>Designing social futures</td>
<td>Civic pluralism</td>
<td>Multilayered lifeworlds</td>
</tr>
<tr>
<td>Fast capitalism/post Fordism</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

One important technical term in multiliteracies is the concept of ‘pedagogy as Design’ (capitalization in the original), which holds that literacy educators and students are supposed to consider themselves as active designers of meanings and social futures (Cope and Kalantzis 2000:19). The foundational work on multiliteracies (New London Group 1996) suggests treating any semiotic activity as a matter of ‘Design’ and conceptualizes Design as embracing three elements: Available Designs (i.e. the grammars of various semiotic systems, orders of discourse, and the intertextual context), Designing (i.e. the semiotic activity of production), and the Redesigned (i.e. the transformed meaning-making resources, or a new Available Design). In doing so, the available semiotic resources, the active process of transformation, and the reproduced new representations are constantly in tension. The three elements together tend to stress the fact that meaning-making is an active and dynamic process instead of something governed by static rules. Based on the conceptual of Design, the New London Group (1996) stress the significance of developing a metalanguage in functional terms to describe the forms of meaning that are represented in Linguistic Design, Visual Design, Audio Design, Gestural Design, Spatial Design and Multimodal Design. Two key concepts, i.e. ‘hybridity’ and ‘intertextuality’
Research Background: Multimodality, Multiliteracies, and EFL Education in China

(Fairclough 1992, 2000), are proposed to describe multimodal meanings and the relations of various types of Designs.

As the New London Group (1996) state, one objective of the multiliteracies schema is to develop a metalanguage to describe meaning in various realms, which include the textual, visual, as well as the multimodal relations between different modes. Their emphasis on the metalanguage echoes what has been pointed out in social semiotics that no single code can be successfully studied or fully understood in isolation, and therefore ‘a theory of verbal language has to be seen in the context of a theory of all sign systems as socially constituted, and treated as social practices’ (Hodge and Kress 1988: vii-viii). The metalanguage is an educationally accessible functional model of language that is needed to support a pedagogy of multiliteracies. According to New London Group (1996), the metalanguage of multiliteracies does not describe the elements of Design as rules, but is a heuristic process that accounts for the infinite variability of different forms of meaning-making in relation to specific cultures and subcultures. As the New London Group state,

> The primary purpose of the metalanguage should be to identify and explain differences between texts, and relate these to the contexts of culture and situation in which they seem to work...In trying to characterize game and genre, we should start from the social context, the institutional location, the social relations of texts and the social practices within which they are embedded. (New London Group 1996)

The emphasis on the strong connection between the meaning-making resources of semiotic systems and their use in social contexts aligns with one of the basic tenets of SFL (Halliday 1973, 1978, 1994; Halliday and Hasan 1985; Martin 1991, 1992; Halliday and Matthiessen 1999, 2004), which accords great significance to the dialectical relationship between linguistic system and the context in which it is embedded. The SFL theoretical rationale that is of special relevance to the present study will be explained in detail in Chapter 3.
The issue on how to put the multiliteracies schema into pedagogy practices seems to be addressed in the multiliteracies manifesto by describing the four-component pedagogy that involves Situated Practice, Overt Instruction, Critical Framing and Transformed Practice (capitalization in the original). As the New London Group (1996) explain, Situated Practice is the immersion aspect that is concerned with the previous and current experiences of the community that consists of both learners and experts. The second element, i.e. Overt Instruction, deals with the explicit teaching of the metalanguage for describing and interpreting Design elements of different modes of meaning. It is hoped that through the collaborative efforts between teacher and student in Overt Instruction, students will be able to accomplish complex tasks that they might not be able to accomplish by themselves. Critical Framing deals with the development of students’ critical and cultural understanding of what they are learning in relation to its context. Through Critical Framing students may become aware of, and thus be able to articulate the cultural connections between practices. Last but not least, Transformed Practice means the transfer or re-production of Design of meanings for the purposes and goals in real-life situations, through which students as ‘designers of social futures’ can put the Redesigned to work in other contexts or cultural sites. It is hoped that students, acting as, can implement their understandings that are acquired through Overt Instruction and Critical Framing into practices in real-life situations. In short, Situated Practice involves recognizing the differences in the economic, civic and personal spheres of life; Overt Instruction is designed to help students develop a metalanguage that accounts for Design differences; Critical Framing centres around associating those Design differences with cultural purposes; Transformed Practice is concerned with moving from one cultural context to another. As has been stressed by the New London Group (1996), it should be noted that the four components of multiliteracies pedagogy do not constitute a linear hierarchy or represent stages. As a matter of fact, the elements of each component may occur simultaneously while at different times one of these elements may predominate, and all of them are repeatedly revisited at different levels.
The four aspects of the multiliteracies pedagogy, as is further clarified by Cope and Kalantzis (2000: 237), actually build on various teaching traditions or approaches. To be specific, Situated Practice draws on the pedagogies of process, immersion and experiential learning; Overt Instruction builds on the pedagogies that explicitly teach rules and conventions; Critical Framing is based on the pedagogies of critique and contextualization; while Transformed Practice builds on the pedagogies that link theory with practice and emphasize the transfer of understanding from one context to another. It is argued that these traditional approaches may be limited when used separately, whereas the integration of the four components will hopefully add considerable depth to each of them. The feasibility of integrating various pedagogical traditions with diverse purposes and practices into one framework, however, has aroused critical comments. Auerbach (2001) fears a possible theoretical ‘distortion and co-option’, and the way to avoid this, as she articulates, is to make explicit ‘the ideological bases of each of the four components and their implications for practice’.

To sum up, recognizing diversities and differences instead of advocating established cultural and linguistic forms, the concept of multiliteracies is designed to supplement traditional literacy pedagogy ‘by addressing two related aspects of textual multiplicity’ (Cope and Kalantzis 2000: 5-6). On the one hand, while traditional practices mainly concentrate on language, the pedagogy of multiliteracies takes into account other channels of communication beyond language. On the other, traditional pedagogies usually focus on a singular standard form of language, whereas the multiliteracies pedagogy argues for the necessity of assisting language learners in describing linguistic and cultural diversity. These two characteristics are also reflected in the major research trends approaching multiliteracies, which will be briefly introduced in the remainder of the section.

2.2.2.2 Other approaches to multiliteracies

When reviewing a further developmental stage of the pedagogy first espoused in the multiliteracies manifesto, Kalantzis and Cope (2001: 9) point out two interrelated
Research Background: Multimodality, Multiliteracies, and EFL Education in China

changes. One is the cultural and linguistic diversity and multiple Englishes, and the other lies in the investigation into the multimodal meaning-making resources (e.g. linguistic, visual, audio, gestural and spatial patterns of meaning). These two trends align with the dual acknowledgement highlighted in the manifesto - namely that multiliteracies research considers not only diversity of cultures and plurality of texts, but also the multimodal modes of meaning that are enabled by modern information and multimedia technologies. Voices that concern themselves with the multimodal channels of communication have been presented in the preceding section. In the present discussion I will briefly outline the other dimension, which concentrates on cultural and linguistic diversity.

The trend in multiliteracies research that centres around linguistic and cultural diversity takes into account what constitutes the appropriate literacy pedagogy in response to the local diversity and global connectedness, as well as the multiple variations of English along with its standard version. The view of language on which multiliteracies is built tends to be a dynamic conception, which places ‘difference, change and creativity’ at its centre (Fairclough 2000). In other words, multilingual literacy researchers argue against the single canonical form of English that is available in literacy education, while endeavouring to validate the diversity of literacies that manifests actual experiences and social changes in the contemporary world.

The intra-lingual diversification that results from globalization has been examined and some paradoxical linguistic consequences have been observed. Lo Bianco (2000) comments that English is becoming a lingua mundi, or world language, and a lingua franca, a common language of global commerce, media and politics. Admittedly, this happens at the expense of linguistic diversity since a large number of small languages are actually facing extinction. However, as Lo Bianco (2000) points out, English is at the same time fracturing into multiple and differentiated Englishes, which are marked by such factors as accent, national origin, subcultural style and technicality. When it comes to literacy teaching and learning, Lo Bianco (2000) argues that multilingualism is an important aspect of multiliteracies, and the mastery
of a single, canonical version of English is far from being sufficient. On the contrary, an educational response to the challenge of language variation and the variation in language attitude is supposed to claim attention in literacy education.

The academics who advocate multiliteracies with the emphasis placed on literacy education explore the institutional and social contexts in developing world as well as post-industrial societies, which deals with not only children literacy but the training of practising teachers. Newfield and Stein (2000) apply the ideas from multiliteracies schema to the institutional context that involves students doing their masters degree in English education in South Africa, where language diversity is perceived by the students as a crucial factor for their future development. The multiliteracies pedagogy is tested in this context both as in pedagogic practice as part of the postgraduate course, and in curriculum documents for critical evaluation by the students who prepare themselves as English teachers. As Newfield and Stein (2000) contend, the survey is designed to function as a critical dialogue between practising teachers and multiliteracies researchers, with the purpose of providing feedback on the relevance of multiliteracies to the specific contexts of actual teaching.

As Kalantzis and Cope (2001: 9) state, the new developments in social and technological realms have the potential to transform the substance as well as the pedagogy of subject English and, in a wider setting, literacy education. Another important task that multiliteracies research has set for itself is to address the cultural diversity that exists in schools as well as in workplaces and community life. Nakata (2000) considers how the multiliteracies pedagogy can be adopted to address the tension between the importance of English literacy and the issue of cultural sensitivity the indigenous context. In order to avoid the mistakes made by colonialism, Nakata (2000) argues that an indigenous standpoint needs to be established by local researchers, which positions indigenous people at the centre of their own life and history rather than as marginalized ‘other’. His discussion tends to reveal that English literacy cannot be bypassed if indigenous students are to deal effectively with the realities of globalization, while the adoption of multiliteracies pedagogy may enable the students to articulate themselves in the current technological and global
order on the basis of an indigenous position.

As was mentioned above, workplace is one of the social domains in which the characteristics of change and transition are evidently demonstrated. Gee (2001) shows us how the ‘socially-situated identities’ in workplace are reflected in schooling. By contrasting teacher expectations of working class students with those of upper middle class students in a culturally diverse elementary school in the United States, he examines the way in which schools are structured to produce different types of workers (i.e. ‘symbol analysts’ at the top, ‘low-level technical workers’ in the middle, and ‘service workers’ at the bottom) for the ‘new capitalism’. His observation is further supported by his interviews with working class and upper middle class teens. As Gee (2001) reveals, working class teens often use narrative language and are ‘immersed in a social, affective, dialogic world of interaction’, whereas the upper class teens tend to use less personalized expressions and are thus ‘immersed in a world of knowledge, argumentation, and achievements’.

Focusing on the academic context, Michaels and Sohmer (2001) investigate students’ mastery of new literacies and the new academic identities that accompany in those intellectual practices in science teaching environment. In investigating the operation of multiliteracies pedagogy in different settings through the collaboration with public school science teachers, they stress the importance of understanding the linkages created by the teacher between the four components of the pedagogy and their distributional patterns. In addition, Michaels and Sohmer (2001) also point out that the concept of Critical Framing demands further description, and suggest involving teachers and non-specialist community members in characterizing and theorizing the multiliteracies pedagogy.

As was reviewed above, the concept of multiliteracies is interpreted in a broad sense to include modes of communication other than language along with literacy education. Owing to the fact that the present thesis focuses on multimodal pedagogic materials for EFL education, the rest of this chapter will be devoted to the linguistic and pedagogic scene in China.
2.3 Multimodality research, textbook studies and curriculum standards in China

As a new research area in linguistics, multimodality has drawn considerable attention within Chinese academic circles in the recent decade. In China, textbook editing and design is closely related to curriculum standards for specific school subjects. This section will delineate the general scenario of the development in these areas, as part of the research background concerning the educational context in China that is relevant to the present study.

2.3.1 Development of multimodality research in China

Multimodality research in China includes both the exploration of fundamental tenets and analytical practices of multimodal texts in various registers. In terms of the introduction of theories on multimodality, Li (2003) provides a survey of the analytical framework for analyzing visual images from social semiotic perspective (Kress and van Leeuwen 1996; van Leeuwen and Jewitt 2001). Moreover, she highlights the significance of multimodality research in English teaching, in that it will deepen people's understanding of language as social semiotic, and argues that the pervasive influence of visual images in many communicative contexts including education calls for systematic description of visual display. Hu (2007) explores the issue of multimodalization in social semiotics (Halliday 1978). Based on the distinction between multimodal semiotics and multimedia semiotics, he points out that multimodal semiotics is based on the view that all texts inherently possess the nature of multimodality. Furthermore, Hu probes into some key issues including computer semiotics, semiotic resources as semiotic system, the replacement of books with screens, and the coherence involving discourse participants. With reference to educational context, Hu emphasizes the significance of nurturing students’ multiliteracy skills in the era marked by multimodal social semiotics. Zhu (2007) points out that the restricted focus on language in discourse analysis has led to the negligence of other ways of meaning-making that happen in actual and natural
communication, and this problem can be solved to a great extent by the research on multimodality. He reflects on four important issues centring around multimodal discourse analysis (Kress and van Leeuwen 2006), i.e. where it comes from, how it is defined, its nature and theoretical basis, and its content, methodology and significance. Further development of multimodality research, as Zhu states, demands cross-disciplinary collaboration, and future research trends in this area include the study on the complementarity between different modes of communication, as well as the increasing complexity (e.g. from two dimensions to three dimensions, from static to dynamic). Zhang (2009) proposes a synthetic theoretical framework for multimodal discourse analysis, which takes into account context of culture, context of situation, the semantic level, the formal level, and the level of media. Furthermore, Zhang examines the relations between different modalities, including complementary and non-complementary relations. Based on an investigation into the ways in which grammatical structures of different modalities realize ideational, interpersonal and textual meanings, he explores how grammatical categories can be established for analyzing non-linguistic modalities.

Systemicists in China have analyzed both static printed page and dynamic multimodal discourse. As for printed media, Ding (2007) examines the modality of fairy tale illustrations from social semiotic perspective. He suggests referring to this type of visual modality as a kind of ‘surrealist illusionism’, which is characterized by entertaining style, simplified perspectives and colouration, condensation of details, absence of background, as well as some slight anatomical distortions. Drawing on Halliday’s (1978: 164) notion of ‘antisociety’, Ding proposes that through reading fairy tale illustrations, children are led to an ‘anti-world’ of their own, which is illusionary and exists symbiotically with the reality. Wang (2007) analyzes how interactive meaning is construed in a poster in the aspects of contact, social distance, angle, as well as modality. This study demonstrates the feasibility and effectiveness of studying advertising discourse by adopting SFL and the analytical framework of visual image (Kress and van Leeuwen 1996) that derives from this linguistic theory. Chen and Wang (2008) study the text-image relations in Chinese geography textbooks
currently used for primary, secondary schooling and university education. Drawing on Martin and Rose’s (2008) conception of ideational meaning construal in technical images and the logical relations between verbiage and image in scientific context, Chen and Wang demonstrate that at the primary level, the ‘iconic’ colourful cartoons and photographs either specify or repeat the verbal report and explanation; as for the secondary textbook, the iconic images with ‘insets’ (Unsworth 2005) either specify or restate what is described in the verbiage on entities and activities; and for the undergraduate science textbook, black-and-white ‘symbolic’ and ‘indexical’ diagrams are adopted for summarizing what is verbally represented in longer texts. The analysis indicates what is assumed by the textbook editors about students’ capacity of reflecting upon different visual styles, and calls for educators’ as well as students’ conscious awareness of the deployment of various semiotic resources in meaning-making.

In the research area of dynamic multimodal texts, Chinese systemicists’ contributions include hypertexts on computer screen, conference presentations that draw on multiple semiotic resources, and multimodal television advertising. Ye (2006) explores the multimodality of hypertexts on http://www.people.com.cn by adopting social semiotic approach. In doing so, he provides a social semiotic account of multimodal texts, and analyzes the hypertexts along three axes, i.e. the interpretation of states of affairs in the represented world, the description of the social relations between viewer and the image, as well as the organization of the visual space (Kress et al. 1997). The third dimension is further divided into three interrelated systems of information value, salience and framing in the composition of visual displays (Kress and van Leeuwen 1996). Ye’s study demonstrates social semiotics as a powerful tool in interpreting and explaining the multimodality in hypertexts. Zeng (2006) investigates the dynamic construal of multimodal semiotics in international academic communication. After identifying the multilayered and overlapping features in the schematic structure of this type of dynamic discourse, Zeng examines the representational function and organizational function (Lemke 1998a) of verbal language, the orientational function of gestures, and the roles of
visual images such as graphs, tables and pictures in academic conference presentations. Her findings indicate that the understanding of multiple semiotic systems is of great importance in learners’ successful multimodal communication. Chen (2008a) examines verbal and visual evaluative resources in television public service advertising for Beijing Olympics, by drawing on APPRAISAL theory (Martin 2000a; Martin and White 2005) and systemic functional semiotic approach to visual images (Kress and van Leeuwen 2006). Her research findings show that inscribed JUDGEMENT and APPRECIATION are dominant in the verbiage of public service advertising. ENGAGEMENT and GRADUATION resources contribute to the alignment with the audience and the enhancement of evaluative stance. In terms of visual displays on television screen, three visual styles (i.e. close-up, visualization of imageries, and visualization of ideational tokens) are identified, which construe evaluation via inscription and invocation respectively. The pattern of multimodal construal of evaluative stance is thus presented to demonstrate the ‘multiplying’ (Lemke 1998a) relationship between verbiage and images in television public service advertising.

It is worth noticing that few, if any, multimodality studies have been carried out within the EFL teaching and learning context in China. This seems even more surprising when it is remembered that great efforts have been made in curriculum design and reform in China’s educational context; to date, however, few studies have addressed the issue of multiliteracies from a linguistic or social semiotic perspective. The following section will review textbook studies and curriculum reform in China, by way of mapping out the educational landscape in which the present study is located.

2.3.2 Textbook studies in China

Textbook editing has always been an important topic in educational reform in China. For instance, the textbooks for primary and secondary education have gone through two main developmental stages, i.e. from a ‘unified compilation system’ to a ‘revision and approval system’. The former refers to the practice involving the Ministry of
Research Background: Multimodality, Multiliteracies, and EFL Education in China

Education promulgating the teaching syllabus and assigning PEP to edit and design textbooks. This practice had lasted more than thirty years from 1950 to the end of the 1980s. The latter practice means that as long as textbook compilation meets the demands of local reality, it is not necessary to adopt the same series of textbooks throughout the country. Various editions including the ‘common edition’, ‘coastal edition’ and ‘inland edition’, have coexisted since the end of the 1980s. The age of ‘one syllabus, multi-edition’ began at the beginning of the 1990s (Wang 2000).

According to The Encyclopedia Americana (2000: 563), in the strict sense, a textbook is defined as ‘a book that presents a body of knowledge in an organized and usually simplified manner for purposes of learning.’ Undoubtedly the textbook nowadays is only one of the members in a large family of teaching materials, which include other printed documents such as exercise books, reference books and maps, as well as multimedia and electronic teaching materials e.g. audio cassettes, video tapes and computer software. However, the textbook is frequently regarded as the most important teaching tool because it determines not only what will be taught but also how it will be taught.

Research on the textbooks for primary and secondary education in China has been carried out at both macro and micro levels (Zhang 2005: 9-11). Macro-studies include those on the positive functions and negative effects of textbook (Wang 2000; Yang 2002), textbook compilation and selection principles (Xie and Song 2003; Yuan and Chen 2007), structure and evaluation of textbook (Ding 2001; Kuang 2002). Studies adopting the micro-approach explore the issues of linguistic sexism (Zhang and Yang 2003), illustration (Chen 2006; Chen and Qin 2007; Chen and Zhang 1997; Cui and Ren 2001; Shen and Tao 2001; Tao and Shen 2003; Wang et al. 2006), and cultural composition and content (Wang 2000; Xiao 2004). In most of the teaching and learning context, communication takes place through the complex interplay of a range of modes that may include language, visual images, and other meaning-making resources (Jewitt 2002; see also Kress et al. 2001; Lemke 2000, 2002). Textbook discourse is no exception, and the role and value of visual images in school textbooks has been recognized and reflected on by some researchers and teachers. For instance,
the empirical studies that compare learning from illustrated text and learning from verbal text alone show that the teaching effect of the former is better than that of the latter (Levie and Lentz 1982). In China, the school subjects that have been covered in the studies on textbook illustrations include history (Du 1982; Wang and Lu 2006; Zhang 1993), biology (Cui and Ren 2001; Jiang 1984), physics (Tang 2007; Zhang 2005), chemistry (Wang et al. 2006; Zuo 1986), Chinese language and literature (Chen 2006; Yan et al. 2003), mathematics (Xu 1997), moral education (Huang 1999), and geography (Bo 2000). However, the majority of these investigations are more like running commentaries by school teachers and textbook reviewers. They are mainly concerned with the ‘How’ (e.g. ‘How to use illustrations in classroom teaching?’ and ‘How to teach the students to understand illustrations?’) but neglect the ‘Why’ (e.g. ‘Why the illustration is needed in this context? Can we do without it?’, ‘Has the illustration achieved what it is supposed to achieve, and why?’, and ‘Is the illustration suitable for the whole class?’) (Chen 2005).

There are some notable exceptions to this phenomenon, which examine the images in textbooks from the psychological and cognitive perspectives. For example, Shen and Tao (2001; see also Tao and Shen 2003) examine the reading comprehension index and eye-movement index for reading texts with or without illustrations by using the Applied Science Lab Model 4200R Eye View Monitoring Systems. Their findings show that both Grade 8 and Grade 11 students have a much better performance in reading texts with illustrations than those without illustrations, which is reflected in the parameters of reading scores, fixation times and regressive times. Their research indicates that visual illustrations can facilitate reading comprehension, especially the reading of simple texts. Song (2005) argues for the three essential functions of textbook illustrations, i.e. decoration, interpretation and facilitation, from the angle of cognitive psychology, and suggests that the criteria for the assessment of textbook illustration should be based on its relevance with the verbal text, its role in the whole text, its presentation style and physical features.

While most studies taking the cognitive and psychological stance have exclusively focused on the role of images in reading comprehension, the functions of
the visual displays in textbook discourse mean more than the cognitive processes and psychological effects. In the new curriculum reform in China, the three-dimensional goal encompassing ‘knowledge and skill, emotions and values, and process and methods’ has been set up, and moral education and values education have been defined from the angles of the integrity and conformity of curriculum functions (Zhong 2006; Zhu 2006). Nevertheless, how the multimodal meaning-making resources in textbooks are co-deployed to embody, or realize the dimension of ‘emotions and values’ in the overall goal is still far less explored. As Zhang (2005: 11-12) states, there remain substantial gaps in the understanding of textbook discourse, which is evident in three aspects of textbook research in China: firstly, few studies have been carried out at the micro-level as compared with those at the macro-level; secondly, how the goal of emotion and attitude education is incorporated in school textbooks remains a topic relatively unexplored; thirdly, the EFL textbooks for primary and secondary education are far less investigated when compared to the textbooks for other school subjects. Being descriptive and explanatory in nature, and with the focus placed on interpersonal management in EFL textbook discourse, the present study will hopefully contribute to the growing body of work in those three areas.

2.3.3 The interpersonal dimension in Curriculum Standards for English

In China, the textbook is most frequently and authoritatively mentioned as ‘the book used for teaching, which is compiled and edited according to teaching syllabus or curriculum standards, and systematically reflects the content of certain subject’ (The Encyclopedia of China 2004: 145).

The close relationship between textbook-editing and curriculum standards calls for an introduction of the current curriculum standards for a better understanding of the textbooks under attention.

The EFL textbooks under discussion are eighteen textbooks for primary education, junior and senior secondary education, which are edited and published by PEP between 2002 and 2006. They belong to the latest edition after the eighth curriculum reform in 2001, whose significance has been highly recognized among
educationists in China. As Zhong (2006) comments on its significance,

Overall Curriculum Design Plan, namely, Outline of Basic Education Curriculum Reform (implementation) (2001) put forward by the Ministry of Education initiated the eighth curriculum reform, and laid out the blueprint of China’s basic education curriculum reform in the new century with a new discourse system... This change is an important milestone in our country’s curriculum development history. (Zhong 2006: 373)

The overall goal of the primary and secondary EFL education, as is stipulated in the Curriculum Standards for English, is an ‘integrated language-using competence’ that encompasses five aspects: language skills, linguistic knowledge, emotions and attitudes, learning strategies, and cultural awareness. As is stated in Curriculum Standards for English, the five aspects synergize to promote the formation of the integrated language-using competence. To be specific, language skills and linguistic knowledge provide the basis for the competence, cultural awareness guarantees the proper use of language, emotions and attitudes are important factors affecting students’ learning and development, and learning strategies assure the development of independent learning (Zhonghuarenmingongheguo jiaoyubu 2001a: 6). Another feature in the current curriculum standards is its adoption of the hierarchical nine-grade system to stipulate the goals in the aforementioned five dimensions (Cheng 2002). When explaining what primary and secondary students are supposed to learn and achieve, the curriculum standards mainly centre around those five aspects in describing the specific objectives and contents.

The curriculum standards for each subject stress the goal of emotion and attitude education, and this is the first time that curriculum standards have had a clear statement on the attitudinal goal of developing students’ positive emotions (Zhu 2006: 193). The Curriculum Standards for English are no exception. For example, in the Curriculum Standards for English (Grades 1 to 9) for Years 3 to 12, the second item
Paying regard to students’ orientation; creating a tolerant, democratic and harmonious atmosphere in teaching:

The students will maintain high spirit and attain achievements only when they develop positive emotions towards themselves, English, western culture, and English learning. Negative emotions will not only affect English learning, but will be harmful to students’ long-term development. Therefore, teachers should constantly pay attention to students’ emotions, and make great efforts to create a tolerant, democratic and harmonious atmosphere while teaching.

(Zhonghuarenmingongheguo jiaoyubu 2001a: 28)

In the *Curriculum Standards for English*, ‘emotions’ and ‘attitudes’ are defined as the interest, motivation, confidence, willpower, cooperation and other relevant factors including national consciousness and international vision that influence students’ learning process and outcome. Positive attitude is seen as a key factor in the success of English learning (Zhonghuarenmingongheguo jiaoyubu 2001a: 21-22). In a nutshell, the significance of emotion and attitude has been recognized in the curriculum standards, and the cultivation of students’ positive emotions, high moral standard and healthy aesthetic interest has become one of the core issues in EFL education. One of the research goals of the present study will be to examine the way in which multimodal resources in EFL textbook discourse are deployed to encode the attitudinal meanings, as well as how these attitudinal meanings are institutionalized and accumulated along the educational ladder from primary to secondary education.

There has been a great deal of discussion concerning the theoretical foundations of the current *Curriculum Standards for English*. Cheng and Gong (2005) maintain that the current *Curriculum Standards for English* takes a functional view of language and it is a type of ‘multi-syllabus’ or ‘mixed syllabus’ that integrates a variety of features of different syllabi. On the other hand, our educationists emphasize that ‘the building of knowledge is not a purely individual business, but a process full of
interactions between the minds of individuals… individuals are influenced by the social cultural environment and construct their own opinions and knowledge by means of direct interaction with other people’; they also acknowledge that ‘the classroom teaching itself is a dialogue process’ which ‘encourages students to have dialogues with the objective world, with other people and with themselves’ (Zhong 2006: 377-378). With its focus on the multimodal meaning-making resources in EFL textbook discourse, the current research will also reflect on how dialogic process is construed by means of multiple semiotic modes, and consider the arrangement of visual styles that manifests textbook editors’ assumption of students’ capacity in handling multimodal texts. As for the latter, attention will be paid to whether a certain visual style is appropriate in relation to some major trends in the twentieth-century educational theory.

2.4 Summary

This chapter has first provided a retrospect of the previous studies on multimodality and multiliteracies, including the introduction of relevant definitions, technical terms and major research trends in these research areas. In terms of multimodality research, emphasis has been placed on the distinction between social semiotic approach and mediated discourse perspective. As for the theorization of multiliteracies, it is noteworthy that the development of a metalanguage in functional terms for the description and explanation of multimodal meaning-making resources has been highlighted, and the strong connection between meaning-making resources and their use in social contexts has been recognized.

The remainder of the chapter has been devoted to the review of multimodality research, studies on textbook, and EFL education scenario in China. This review has identified five research areas that demand further exploration. First, the multimodal features in China’s pedagogic context are far less understood than those in the west. Second, EFL teaching and learning resources are much less investigated in comparison with those for other school subjects. Third, while dialogic process is
advocated for classroom teaching and textbook is recognized as an essential component in this process (e.g. Chen and Ye 2006), the ways in which multimodal resources are deployed to manage heteroglossic space in textbooks has been undertheorized. Fourth, for the first time ‘emotion and attitude education’ has been highlighted in the reformed curriculum standards. However, how linguistic and visual meaning-making resources in textbooks are deployed to realize the ‘emotion and attitude goal’ has yet to be examined. Fifth, image producers’ assumptions about the viewers, as reflected in the choice of visual arrangement in textbooks, call for a semiotic explanation.

To avoid too many foci, the present thesis will analyze and discuss the verbal texts and visual images in EFL textbook discourse among many other multimodal teaching and learning resources, and focus on the interpersonal meaning these multimodal meaning-making resources are deployed to realize. Three major research gaps will be further specified and suggestions will be put forward in Chapter 3 in relation to the theoretical foundations that underpin the current research.

NOTES

1 The term ‘Sydney School’ is proposed by Green and Lee (1994), but in fact the studies carried out by the functional linguists and educational linguists in the Department of Linguistics at the University of Sydney have influenced work around and well beyond Australia, reaching Indonesia, Singapore, Hong Kong, Britain, Argentina, South Africa and many other places in the world. Martin (2000b) uses the term ‘action research’ to refer to this work because it ‘involved an interaction of theory and practice which pushed the envelop of the understandings about modelling language in social life and which at the same time led to innovative literacy teaching across sectors in Australia and overseas’.


3 In the educational system in China, Grades 1-6 refer to the six years of primary education; junior secondary education is referred to as Grades 7-9 and senior secondary education Grades 10-12. Therefore, Grade 8 means the second grade in junior secondary school, and Grade 11 is the second grade in senior secondary school.

4 The terms ‘curriculum standards’ and ‘teaching syllabus’ have been used alternatively in the history of education in China. The first curriculum standards were promulgated in 1912 and were used until 1952. They were then replaced by teaching syllabus under the influence of Soviet pedagogy theories (Liao 2004). The contemporary sense of ‘curriculum standards’ is the product of the standards movement during the 1990s (Richards 2001). Its function is quite similar to that of a teaching syllabus, and in many countries including China, the teaching syllabus has been replaced by curriculum standards (see Cheng and Gong 2005).

5 Generally speaking, Grade 2 in the nine-grade system refers to the English level of Year 6 students who start to learn English at Year 3; Grade 5 means the English level of Year 9 students while Grade 8 refers to the English level of senior secondary school graduates. In other words, Grades 2, 5, and 8 are the general English levels for primary, junior and senior secondary school graduates respectively. However, the relationship between ‘Grade’ and ‘Year’ is by no means a direct one. Different regions and schools can modify the goals for the English curriculum according to the local realities (Zhonghuarenmingongheguo jiaoyubu 2001: 4-5; see also Cheng 2002).

6 There are three *Curriculum Standards for English*: the first one was published in 2001, which is for Grades 1 to 12 covering both the nine-year compulsory education and senior secondary education; the second one was published in 2003 and focuses only on senior secondary English; the third one was published in 2005 and concentrates on the nine-year compulsory education. These three curriculum standards have a common theoretical foundation, and the latter two are based on the first curriculum standards (Cheng and Gong 2005).
Chapter 3

Theoretical Foundations

3.1 Introduction

As was mentioned in Chapter 1, the main theoretical foundation for the present study will be SFL, particularly its recent advances in interpersonal semantics and the systemic functional semiotic approach to visual images. Taking an inter-organism perspective and attaching great significance to language functions, SFL has found wide applications in both linguistic and semiotic fields of studies. The purpose of this chapter is two-fold: firstly, some of the major theoretical assumptions of SFL that are closely related to the current research will be introduced and explained; and secondly, the rationale of an SFL approach to the present multimodal pedagogic materials will be sketched before looking in detail at specific examples in the ensuing chapters. As for the explanation of the theoretical framework, I will draw on the pioneer work of Halliday (1973, 1978, 1979/2002, 1994; Halliday and Matthiessen 1999, 2004) as well as the interpretations and recent developments by other systemicists and social semioticians (e.g. Eggins 2004; Hu et al. 2005; Kress and van Leeuwen 1996, 2001, 2006; Lemke 2002; Martin 1985a, 1992, 1997, 1999a, 2000a, 2001, 2004; Martin and Rose 2007a; Martin and White 2005; Thompson 1996).

3.2 Some basic tenets of SFL

Taking an inter-organism perspective to study language in relation to its social context, SFL centres around language function, not only explaining the functions of language
but also exploring how language is structured according to its functions (Halliday 1973, 1978). The present section is devoted to a brief introduction to the relevant theoretical preliminaries and rationales of SFL, which regards language use as ‘functional, semantic, contextual and semiotic’ (Eggins 2004: 3).

### 3.2.1 Language as social semiotic

The functional orientation of SFL attaches great significance to the social context in which verbal or any other semiotic communication is embedded. ‘Language as social semiotic’ (Halliday 1978) reveals the strong connection between meaning-making practices and the relevant social and cultural contexts. The meanings of semiotic choices depend on the social context of use. As Halliday points out,

> … the formulation ‘language as social semiotic’… means interpreting language within a sociocultural context, in which the culture itself is interpreted in semiotic terms – as an information system. (Halliday 1978: 2)

The conception of ‘language as social semiotic’ is often referred to as the common interest to all systemic linguists despite their different research emphases or application contexts (Eggins 2004: 3). As is interpreted by Lemke (2002), in social semiotics language use is considered to be one of the semiotic practices that makes use of the semiotic resource system of language to make social meanings. Therefore, it can be inferred that the approach of social semiotics is in principle able to show how social meanings and social relations are constructed by the deployment of other semiotic meaning-making resources beyond language.

In explaining the ‘natural dialectic’ relation between language and social system, Halliday (1978: 183-192) distinguishes between ‘language as system’ and ‘language as institution’ when presenting the underlying conceptual framework. As Halliday (1978: 183) states, ‘the salient facts about language as system are (a) that it is stratified (it is a three-level coding system consisting of semantics, lexicogrammar
and phonology) and (b) that its semantic system is organized into functional components (ideational, including experiential and logical; interpersonal; textual). Language as *institution* indicates the fact that language is *variable*, and there are two kinds of variation, (a) *dialect* (variation according to the *user*), and (b) *register* (variation according to the *use*)’ (emphasis in the original). The rest of the present section will consider these basic tenets. In terms of ‘language as institution’, the focus will be placed on the registerial variation in relation to context of situation due to its high relevance to the present study.

### 3.2.2 Stratification within and between semiotics

Drawing on Hjelmslev (1961) in making a distinction between the linguistic planes of content and expression, SFL takes a further step and divides the content plane into the strata of semantics and lexicogrammar, with a natural rather than arbitrary relation between them. This articulation of wording as naturally related to meaning reflects one of SFL’s key assumptions - namely that that language is structured in light of its functions. The relation between the content plane and expression plane (i.e. the stratum of sounds or graphology) is arguably an arbitrary and conventional one in Saussure’s sense. In other words, language in the SFL model is theorized as a tri-stratal (i.e. semantics, lexicogrammar, and graphology/phonology) semiotic system, with a relationship of realization between every two adjoining strata. The systemic model of the strata of language, including both everyday and technical terms, is presented in Table 3-1. The oblique downward arrows signals that the higher stratum is realized by the lower one. In the discussion that follows, I shall concentrate mainly on the strata of semantics and lexicogrammar within the semiotic system of language.
In SFL, semantics operates at the highest stratum within the linguistic system and thus serves as the interface with the context in which the semiotic system of language is embedded. As Halliday (1978: 135) points out, the semantic unit of a text ‘is to the semantic system what a clause is to the lexicogrammatical system and a syllable to the phonological system’. Therefore, the stratum of semantics is also referred to as discourse semantics in that it ‘focuses on text-size rather than clause-size meanings’ (Martin 1992:1). Three strands of meanings (i.e. ideational, interpersonal, and textual) are simultaneously construed at this stratum (see Section 3.2.4 below), and this functional diversification enables discourse analysts upwardly to investigate the linguistic responses to the diversified contextual demands, and downwardly to examine the corresponding lexicogrammatical realizations. The following discussion will consider these two neighbouring strata to discourse semantics.

The stratum of lexicogrammar, which is naturally related to the discourse semantics stratum and part of the content plane, encompasses realizations of discourse semantics. Lexicogrammatical descriptions take the clause as the basic unit of analysis. The system of transitivity, mood structure, thematic structure and other structural configurations by which the clause is organized as representation, exchange and message are identified to capture the lexicogrammatical realizations of the three strands of simultaneously-construed meanings (see Halliday 1994; Halliday and Matthiessen 2004).

As was mentioned above, in SFL language and language functions are examined with regard to the social context in which the language system is embedded.
According to Halliday (1978: 109-110, 122-123), social context is a semiotic structure or semiotic environment in which communication takes place. The relation of language system to the higher-level social context is not merely one of expression, but a ‘more complex natural dialectic’ one, in which ‘language actively symbolizes the social system, thus creating as well as being created by it’ (Halliday 1978:183). Halliday and Matthiessen’s (2004) conception of the stratification of language in context is summarized in Figure 3-1 below.

![Figure 3-1 Stratification (Halliday and Matthiessen 2004: 25)](image)

In accounting for social context, the present study draws on the stratified model developed by Martin (1985a, 1992, 1999a), whose advantages have been argued on both theoretical and practical grounds (Hasan and Martin 1989; Painter and Martin 1986). In this multi-stratal model, context is interpreted as consisting of three communicative planes, i.e. register, genre, and ideology (Martin 1986, 1992: 501-508). Drawing on Hjelmslev’s (1961) work on connotative semiotic system that has another denotative semiotic system as its expression plane, Martin proposes (1992: 495-496) that language functions as the expression form of register (context of situation), which in turn functions as the expression form of genre (context of culture). Genre, register and language function as the expression form of ideology, which is the fourth
communicative plane that has the highest level of abstraction. This stratified model of context is represented in Figure 3-2 below.

![Figure 3-2 Language and its semiotic environment (Martin 1992: 496)](image)

Later discussions on context (e.g. Martin 1997, 1999a) have primarily focused on the strata of register and genre. In this model, the use of the term register is extended to cover both the context’s expression plane, i.e. the linguistic meanings in a given situation type, and the context’s content plane. In other words, the notion of register is used for referring to the semiotic system constituted by the variables field, tenor, and mode, which reflects the metafunctional diversification of language projected onto social context. As is explained by Martin,

… in the model developed here, register is the name of the metafunctionally organized connotative semiotic between language and genre. This means that instead of characterizing context of situation as potential and register as (context’s) actual, … [register is treated] as a semiotic system in its own right, involving notions of both system and process. (Martin 1992: 502)

Instead of deeming genre as one dimension of mode (cf. Halliday 1978: 62), the
stratified model considers genre to be a stratum of higher-level abstraction distinguished from register, which is defined as ‘staged, goal-oriented social process’ (Martin 1986); formulated in these terms, genre takes responsibility for the integration of the semantic diversification that is projected from the functionally organized language system onto the stratum of register. Thus, a metafunctionally diversified view on context is presented through register, whereas genre provides a standpoint from which analysts can transcend any metafunction to consider a text’s global social purpose and the relations between genres.

As for the interstratal relationship, Lemke’s (1984: 35-39; 1995: 104, 168-169) notion of ‘metaredundancy’ is employed to account for the relations of genre to register and language. Genre ‘metaredounds’ with register, which in turn ‘metaredounds’ with language. To put it more simply, social context is modelled as a pattern of linguistic patterns. As Halliday and Martin (1993: 42) point out, the concept of metaredundancy is directional as far as degrees of abstraction is concerned, but not directional as far as cause and effect are concerned; another advantage of this concept is that it is probabilistic rather than categorical, which allows one level of abstraction to reconstrue another and thus the system may evolve. It should be noted that the probabilistic realization relationship between context and language in this model is the same with what Halliday articulates about the realization relationship between field, tenor and mode and language, and yet it is different from the relations between strata within language (Martin 1992, 1999a). Martin (1999a) also suggests a proportionality between the realization of genre in register and the realization of discourse semantics in lexiogrammar, in the sense that genre and register offer complementary perspectives on context just as discourse semantics and lexicogrammar provide complementary facets of the linguistic meaning-making resources. The different types of realization across strata and the aforementioned proportionality can be visually summarized in the ‘stepped box’ diagram (see Figure 3-3 below), with two-headed arrows to represent the two-way dialectic of realization.
3.2.3 Realization and instantiation

So far the focus here has been on the inter-stratal relationship, i.e. realization, which has drawn much attention in SFL research (e.g. Halliday 1994; Halliday and Matthiessen 2004; Martin 1992; Matthiessen 1993). Another kind of relationship in the theoretical architecture of SFL that equally deserves our attention is instantiation⁵, which theorizes the relationship of system to text as the generalized potential to the actual instance that is enacted from the potential. As was reviewed in Section 3.2.3, realization is a scale of abstraction, whereas instantiation on the other hand is a scale of generalization. A well-known analogy of climate and weather provided by Halliday (1999) is often drawn upon in explaining the hierarchical relationship of instantiation:

Climate and weather are not two different things; they are the same thing that we call weather when we are looking at it close up, and climate when we are looking at it from a distance. The weather goes on around us all the time; it is the actual instances of temperature and precipitation and air movement that you can see and hear and feel. The climate is the potential that lies behind all these things; it is the weather seen from a distance by an observer standing some way off in time. (Halliday 1999, emphasis in the original)
It can be inferred from the analogy that system and instance are the same phenomenon viewed from different ends of a continuum. Along this ‘cline of instantiation’ (Halliday and Matthiessen 1999: 385), we may look at language as a generalized system of meaning making potential at one end, and identify the actual reading, i.e. the ‘act of reader/listener interpretation’ at the other end of the cline (Martin and White 2005:162-163). Figure 3-4 summarizes the scale from the language system to the final reading of individual texts as follows.

![Figure 3-4 Cline of instantiation (Martin and White 2005: 25)](image)

Exploring the relatively undertheorized research area of instantiation, Martin (2008b) proposes two dimensions of analysis that call for further investigation: coupling and commitment. Coupling refers to the ways in which meanings combine as a number of coordinated choices from system networks. Some initial work in this area has been done by Nesbitt and Plum (1988) in their study of clause complexing. As is pointed out by Martin (2008b), this orientation to instantiation can be and should be further pursued, for instance, from the perspective of the discourse semantics of APPRAISAL (Martin and White 2005). The options from the simultaneous sub-systems within APPRAISAL (i.e. ATTITUDE, ENGAGEMENT and GRADUATION) can be coupled, and the logic of system networks enables discourse analysts to identify the relevant coupling at any point in delicacy. Furthermore, coupling can be across simultaneous systems not only within a metafunction but between metafunctions. With the interactions at the discourse semantic level between metafunctions, corresponding lexicogrammatical realizations are deployed across clause and group ranks. Therefore, coupling can also be identified across strata and ranks. Crucially, ‘What realization freely combines, instantiation
specifically couples’ (Martin 2008b). The hierarchy of instantiation constrains what can be combined so as to make it recognizable as belonging to a certain generalized, conventional combinations in a given culture. The concept of coupling brings out the importance of the combinations of meaning across systems.

Another orientation to instantiation that I shall draw on in the analyses below is commitment, which is concerned with the amount of meaning potential activated or manifested in a particular process of instantiation. Commitment in other words, deals with the degree to which meanings are taken up and the degree of delicacy selected (Martin 2008b). In terms of ideational meanings, Hood (in press) reveals that a great variety of relations, including de/classification, de/composition, role/incumbent identification, de/specification, metaphor/congruence, and defusion/infusion, offer space for committing different degrees of meaning potential. As for the interpersonal semantics under attention, particularly the APPRAISAL system, the concept of commitment can be adopted to re-interpret the various degrees of explicitness in encoding attitudinal meanings, i.e. the cline from overt inscription to the covert ‘flagging’ with non-core vocabularies and ‘affording’ with ideational selections (Martin and White 2005: 61-67). As Martin (2008b) points out, in terms of commitment, direct inscription is more attitudinally committed than flagging, which in turn is more committed than affording.

3.2.4 Metafunctional diversification

Another aspect within SFL that I am going to review in the present section is the three metafunctions, around which the semiotic system of language and social context are organized. The notion of function is accorded a special status in SFL, in that it is interpreted as ‘a fundamental property of language itself, something that is basic to the evolution of the semantic system’ (Halliday and Hasan 1985: 17). I shall first provide a general account of the metafunctionally diversified semantic components with their lexicogrammatical realizations, and then consider the functional diversification of social context. Due to the focus of the present thesis, particular emphasis will be placed on the interpersonal metafunction.
As is formulated by Halliday (1994: xiii), all languages are organized around ideational and interpersonal meanings, which manifest ‘the two very general purposes which underlie all uses of language: (i) to understand the environment (ideational), and (ii) to act on the others in it (interpersonal).’ The third aspect in this semantic complex is textual meaning, which is intrinsic to the language system and ‘breathes relevance into’ (Halliday 1994: xiii) the other two metafunctions. As is pointed out by Halliday (1992), the concept of metafunction is a theoretical category, which is general to all languages and has a fundamental place in the construction of a general linguistic theory.

The aforementioned three strands of meanings are simultaneously encoded in almost every utterance of verbal communication. The ideational metafunction, which involves the speaker as an ‘observer’ (Halliday 1978: 112), represents not only the processes in the outside world and in the individual consciousness, but also the relations between these processes (Halliday and Matthiessen 1999: 511). They are accordingly categorized into experiential metafunction and logical metafunction respectively. The corresponding functional configuration that is projected onto lexicogrammar makes up the strand of ‘clause as representation’ in the overall meaning of a clause, involving transitivity systems as the lexicogrammatical realization of experiential metafunction, and logical independency relations and logico-semantic relations in the clause complex as the lexicogrammatical resources for logical metafunction (Halliday 1994; Halliday and Matthiessen 2004). The theoretical sketch here is brief and more detailed descriptions will be provided where required in the course of data analysis.

The interpersonal metafunction ‘represents the speaker’s meaning potential as an intruder’, and it is the ‘participatory function of language’ (Halliday 1978: 112). It deals with the construction of social relationships, whereby language constructs our social collective as well as personal being (Halliday and Matthiessen 1999: 511). In other words, interpersonal metafunction is concerned with how people use language to interact with each other, to enact and maintain social relations, to express one’s own attitudes, and to influence others’ opinions and behaviours. Systemicists have
Theoretical Foundations

approached various kinds of interpersonal meaning under the headings of mood, modality, evaluation and negotiation (Thompson 1996: 39); see also Eggins 2004; Halliday 1994; Halliday and Matthiessen 2004; Martin 1992; Martin and Rose 2007a). The diverse aspects of interpersonal management may be summarized into two broad categories, i.e. interactive and personal (e.g. Thompson 1996: 69). The former deals with the interactions between communicating parties, whereas the latter is concerned with personal articulations of evaluation and modality judgement.

Recent developments in SFL have modelled the interpersonal semantics as a set of three broadly-based systems: APPRAISAL, NEGOTIATION, and INVOLVEMENT (Martin 1997; Martin and White 2005: 34-35). NEGOTIATION is concerned with interaction as an exchange between speakers, including the way speakers adopt and assign roles to one another in dialogue and the organization of moves. It involves mood-based resources for exchanging information and goods-and-services (Martin 1997; Martin and Rose 2007a: 219). INVOLVEMENT focuses on non-gradable resources for solidarity. It refers to meanings by which interlocutors code social closeness arising either from intimacy or from their shared membership in some distinct discourse community (Martin and White 2005: 33-34, White 1998). APPRAISAL is concerned with ‘evaluation – the kinds of attitudes that are negotiated in a text, the strength of the feelings involved and the ways in which values are sourced and reader aligned’ (Martin and Rose 2007a: 25). A detailed account of the APPRAISAL system (Martin 1997; Martin and White 2005) that informs the present study will be provided in Section 3.3. Considering the wide application of APPRAISAL to verbal texts in a great varieties of registers (e.g. Coffin 2000; Eggins and Slade 1997; Iedema et al. 1994; Macken-Horarik 2003; Rothery and Stenglin 2000; White 2003), the analysis and discussion of the multiple semiotic resources that encode APPRAISAL meanings remain relatively undeveloped within SFL, which gives impetus to part of the goals of the current study.

The textual metafunction is the ‘relevance function’ of language in the creation of text. It is the enabling functional component that allows the other two metafunctions to operate, and construes the relation of language to its verbal as well
as situational environment (Halliday 1978: 48, 112-113; Halliday and Matthiessen 1999: 512). According to Halliday and Matthiessen (2004: 579), there are two kinds of lexicogrammatical resources that realize the textual meanings in English: structural resources, including thematic structure and information structure that operate at the level of clause, and cohesive resources like reference and conjunction that make explicit the relations between clauses and clause complexes. These two types of resources work together in the marking of points of textual transition as well as textual status.

When considering the relationship of the three modes of meaning to the types of structure at clause level, Halliday (1979/2002: 209) explains that experiential structures tend to be elemental, interpersonal structures tend to be prosodic and textual structures tend to be cumulative or periodic. In extending the discussion of the association between structural types and modes of meanings from clause to text, Martin (1997) suggests a complementarity of particulate, prosodic and periodic structure: particulate structure organizes text segmentally into orbital or serial patterns, depending on whether there is one nuclear segment; prosodic structure is supra-segmental and spreads itself across a text; periodic structure is wave-like and it organizes a text into a rhythm of peaks and troughs. As Halliday (1979/2002: 205) points out, interpersonal meaning is ‘strung throughout the clause as a continuous motif or colouring’ and ‘the effect is cumulative’; this type of realization is referred to as ‘prosodic’ in that ‘the meaning is distributed like a prosody throughout a continuous stretch of discourse’. Martin and White (2005: 19-21) further identify three types of prosodic realization, i.e. saturation, intensification, and domination. Periodic structure is wave-like, organizing a text into a rhythm of highs and lows in the service of information flow. The association between modes of meaning and the corresponding structuring principles is summarized in Figure 3-5 below.
In SFL, the intrinsic metafunctional diversification within language can be upwardly projected onto social context. In other words, the organization of language and social context are both functionally diversified along similar lines, and therefore there is a proportionality between the three metafunctions, i.e. ideational, interpersonal, and textual components, and the three register variables, i.e. field, tenor, and mode. This hook-up is represented in Figure 3-6. Field refers to the social action or institutional practice; tenor focuses on the set of social relations between relevant participants; and mode is the effect of the communicative channel or medium on texture. These three variables together constitute the metafunctionally organized connotative semiotic system of register. As was explained in Section 3.2.2 above, the realization relationship between register and language is probabilistic and dialectic, in the sense that language construes, is construed by, reconstrues and is reconstrued by social context (Martin 1999a). Both intrinsic functionality as well as extrinsic functionality are outlined in detail by Halliday (1978), Halliday and Hasan (1985), and Martin (1991, 1992, 1999a).
3.3 The semantic system of APPRAISAL within SFL

As was noted above, one of the theoretical underpinnings in the present study is the model of APPRAISAL system (Martin 1997, 2000a; Martin and Rose 2007a; Martin and White 2005) within the broader theoretical framework of SFL (Halliday 1994; Martin 1992; Halliday and Matthiessen 1999, 2004). Taking a functional view to model interpersonal meaning at the level of discourse semantics, the APPRAISAL system is designed as encompassing three sub-systems (i.e. ATTITUDE, ENGAGEMENT, and GRADUATION) for exploring the attitudinal meanings encoded in a text, the ways in which space can be opened up or closed down for different voices, and the resources for manipulating the strength of feelings or degree of alignment. The APPRAISAL system with its sub-systems is outlined in Figure 3-7 below, with the square brackets representing either/or choices while the curly brackets representing both/and choices. Each of these sub-systems has its own sub-categories or options, and all these options within the APPRAISAL system are semantic ones that transcend diverse lexicogrammatical structures (Hood 2004: 13-14; see also Martin and Rose 2007a; Martin and White 2005). In this thesis it will be suggested that a wide range of meaning-making resources, including those that are non-verbal, can be brought together and considered systematically as encoding common evaluative meanings and
constructing a global rhetorical orientation. In this section some basic concepts of the APPRAISAL system will be introduced and its key features will be explained, while necessary elaborations will be provided in Chapters 4 and 5, which are devoted to the analyses of attitudes and voices in the multimodal texts in our data.

3.3.1 ATTITUDE
As was demonstrated in the overview of APPRAISAL system above, ATTITUDE is one of the three interacting dimensions; it is concerned with emotional responses, judgements of human behaviours, and evaluations of products and processes. ATTITUDE can be subdivided into three semantic regions – AFFECT, JUDGEMENT, and APPRECIATION (which have been referred to as emotions, ethics, and aesthetics in traditional parlance). In terms of verbal communication, ATTITUDE is realized through attitudinal lexis (e.g. people are grateful because their food has been gathered for the winter, he said I was hard-working, my favorite subject is science), and mental process of reaction (e.g. we love animals, people were shocked by that episode).
Along with overt realizations, ATTITUDE can also be indirectly invoked through various strategies. For instance, there is no explicit attitudinal meanings inscribed in the clause *everywhere they looked nearly everything was destroyed*. However a sense of fear may be invoked by the scene of complete devastation described in the clause.

AFFECT is considered as situated at the heart in the semantic system of ATTITUDE, since it concerns inborn feelings from which the institutionalized JUDGEMENT and APPRECIATION later develop (Martin 2000a; Martin and White 2005: 42; Painter 2003). Martin (2000a; see also Martin and White 2005: 46-52) identifies the following six factors in classifying AFFECT:

-- whether the emotions are construed as positive or negative (e.g. *satisfied* vs. *upset*);
-- whether they are realized as an emotional surge or as an internal disposition (e.g. *attentive* vs. *satisfied*);
-- whether the feelings are reactions to a certain emotional stimulus or as an undirected ongoing mood (e.g. she *disliked* swimming vs. she was *upset*);
-- whether the emotions involve an reaction to an existing stimulus or an intention towards a prospective stimulus (e.g. she *liked* meeting new friends vs. she *looked forward to* meeting new friends);
-- whether the emotions are concerned with un/happiness (e.g. *misery/cheer*), in/security (e.g. *nervous/confident*), or dis/satisfaction (e.g. *bored with/satisfied*); and -- the degree of the intensity of emotions (e.g. from low through medium to high values: *dislike, hate, loathe*).

More specific information with respect to types of AFFECT will be provided in Section 5.3.1.1 below.

JUDGEMENT is concerned with the evaluative meanings, either positive or negative, that construe the attitudinal positioning towards human characters and behaviours. According to Martin (1997: 23; see also Martin 2001, 2004),
JUDGEMENT can be regarded as the institutionalization of feelings in the context of proposals. In other words, JUDGEMENT recontextualizes individual emotions in the realm of proposals about human behaviour; these proposals may be formalized ideationally as rules and regulations in a given culture and become part of field. The attitudinal sub-system of JUDGEMENT is generally divided into two broad categories, i.e. ‘social esteem’ and ‘social sanction’, with five sub-types. JUDGEMENT of social esteem involves moral rules; sub-types included in this category are ‘normality’ (how usual someone is), ‘capacity’ (how capable someone is) and ‘tenacity’ (how resolute someone is). In contrast, JUDGEMENT of social sanction has legal implications, and this category deals with ‘veracity’ (how truthful someone is) and ‘propriety’ (how ethical someone is).

The attitudinal system of APPRECIATION encompasses evaluations of either man-made or natural, concrete or abstract phenomena and processes. These evaluative meanings are discussed under the headings of ‘reaction’ (the product or process is evaluated in terms of its impact or quality), ‘composition’ (the product or process is evaluated according to its complexity or conformity to organizational principles) and ‘valuation’ (the product or process is evaluated based on its social value). Like AFFECT and JUDGEMENT, positive as well as negative evaluations can be identified in APPRECIATION. As with AFFECT, JUDGEMENT and APPRECIATION can be directly inscribed in discourse through the use of attitudinal lexis, or through other indirect means such as the selection of ideational meanings to invoke evaluations of people and things (Martin 2000a; see also Martin and White 2005: 61-68). I shall elaborate on the issue of attitudinal inscription and invocation in Chapter 5 below, when visual as well as verbal meaning-making resources are discussed.

3.3.2 ENGAGEMENT

The sub-system of ENGAGEMENT within APPRAISAL draws on Bakhtin’s (1981, 1986) notions of dialogism and heteroglossia; it comprises networks of options for opening up or closing down the heteroglossic space for voices in a text. ENGAGEMENT resources cover a wide range of devices that construe a heteroglossic communicative
setting for a text, including alternative points of views and anticipatory responses from the audience.

The ENGAGEMENT network covers both monoglossic and heteroglossic aspects. The focus of this study is mainly on how multiple voices are mediated through multimodal meaning-making resources, and discussion here is thus concentrated on the heteroglossic dimension. According to Martin and White (2005: 94), the ENGAGEMENT network covers ‘all those locutions which provide the means for the authorial voice to position itself with respect to, and hence to “engage” with, the other voices and alternative positions construed as being in play in the current communicative context’. As Martin and White (2005) point out, the taxonomy of ENGAGEMENT meanings include four main categories:

[disclaim]: the textual voice positions itself as at odds with, or rejecting some contrary position;

[proclaim]: by representing the proposition as highly warrantable (compelling, valid, plausible, well-founded, generally agreed, reliable, etc.), the textual voice sets itself against, suppresses or rules out alternative positions;

[entertain]: by explicitly presenting the proposition as grounded in its own contingent, individual subjectivity, the authorial voice represents the proposition as but one of a range of possible positions – it thereby entertains or invokes these dialogic alternatives;

[attribute]: by representing proposition as grounded in the subjectivity of an external voice, the textual voice represents the proposition as but one of a range of possible positions – it thereby entertains or invokes these dialogic alternatives (Martin and White 2005: 97-98)

More delicately, these ENGAGEMENT meanings can be divided into sub-categories (e.g. [Disclaim] is subdivided into [deny] and [counter].) to specify dialogic positioning; less delicately they can be generalized into two broad categories of ‘dialogic expansion’ and ‘dialogic contraction’ in light of whether they are opening up
or closing down the heteroglossic space. Those resources under the headings of [disclaim] and [proclaim] are dialogically contractive, while those under the categories of [entertain] and [attribute] are dialogically expansive. Figure 3-8 below outlines the ENGAGEMENT system with typical realizations at the level of lexicogrammar.

![Diagram of ENGAGEMENT system]

**Figure 3-8** The ENGAGEMENT system (Martin and White 2005: 134)

The following chapter will focus on the analysis and discussion of ENGAGEMENT system; multimodal ENGAGEMENT devices will be identified and their contributions to the negotiation of the heteroglossic diversity in EFL textbook discourse will be explored.

### 3.3.3 Graduation

Another sub-system within APPRAISAL is GRADUATION, which accommodates meaning-making resources for scaling attitudinal meanings and ENGAGEMENT values.
Theoretical Foundations

As is pointed out by Martin and White (2005: 135-136), gradability is ‘a defining property of all attitudinal meanings’ in that all the values of AFFECT, JUDGEMENT and APPRECIATION construe higher or lower degrees of positivity or negativity, as is demonstrated in Table 3-2 below with lexicogrammatical realizations. Gradability is also ‘generally a feature of the ENGAGEMENT system’ because ENGAGEMENT values scale for the degree of the speaker/writer’s investment in a given value position, which is represented in Table 3-3. In other words, GRADUATION is central to the whole system of APPRAISAL, and ATTITUDE and ENGAGEMENT can be regarded as domains of GRADUATION that differ only in terms of the nature of the meanings being scaled (Martin and White 2005: 136).

**Table 3-2** The gradability of attitudinal meanings (Martin and White 2005: 136)

<table>
<thead>
<tr>
<th>JUDGEMENT</th>
<th>low degree</th>
<th>high degree</th>
</tr>
</thead>
<tbody>
<tr>
<td>player</td>
<td>competent player</td>
<td>good player</td>
</tr>
<tr>
<td></td>
<td>reasonably good</td>
<td>quite good player</td>
</tr>
<tr>
<td></td>
<td>player</td>
<td>very good player</td>
</tr>
<tr>
<td></td>
<td></td>
<td>extremely good</td>
</tr>
<tr>
<td></td>
<td></td>
<td>player</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>AFFECT</th>
<th>low degree</th>
<th>high degree</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>contentedly</td>
<td>happily</td>
</tr>
<tr>
<td></td>
<td>slightly upset</td>
<td>somewhat upset</td>
</tr>
<tr>
<td></td>
<td></td>
<td>very upset</td>
</tr>
<tr>
<td></td>
<td></td>
<td>ecstatically</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>APPRECIATION</th>
<th>low degree</th>
<th>high degree</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>a bit untidy</td>
<td>somewhat untidy</td>
</tr>
<tr>
<td></td>
<td>attractive</td>
<td>very untidy</td>
</tr>
<tr>
<td></td>
<td></td>
<td>completely untidy</td>
</tr>
</tbody>
</table>

**Table 3-3** The gradability of ENGAGEMENT values (Martin and White 2005: 136)

<table>
<thead>
<tr>
<th>lower</th>
<th>higher</th>
</tr>
</thead>
<tbody>
<tr>
<td>[entertain]</td>
<td></td>
</tr>
<tr>
<td>I suspect she</td>
<td>I believe she</td>
</tr>
<tr>
<td>betrayed us</td>
<td>betrayed us</td>
</tr>
<tr>
<td>possibly she</td>
<td>probably she</td>
</tr>
<tr>
<td>betrayed us</td>
<td>betrayed us</td>
</tr>
<tr>
<td>she just possibly</td>
<td>she possibly</td>
</tr>
<tr>
<td>betrayed us</td>
<td>betrayed us</td>
</tr>
</tbody>
</table>

| [attribute]        |                             |
| she suggested that | she stated that             |
| I had cheated      | I had cheated               |
|                   | she insisted that           |
|                   | I had cheated               |
Theoretical Foundations

<table>
<thead>
<tr>
<th>[pronounce]</th>
<th>I’d say he’s the man for the job</th>
<th>I contend he’s the man for the job</th>
<th>I insist that he’s the man for the job</th>
</tr>
</thead>
<tbody>
<tr>
<td>[concur]</td>
<td>admittedly he’s technically proficient (but he doesn’t play with feeling)</td>
<td>certainly he’s technically proficient (but ….)</td>
<td></td>
</tr>
<tr>
<td>[disclaim]</td>
<td>I didn’t hurt him</td>
<td>I never hurt him</td>
<td></td>
</tr>
</tbody>
</table>

As is shown in Figure 3-7 above there are two axes along which the semantic system of GRADUATION operates, i.e. FORCE and FOCUS. The former grades meanings in terms of the intensity or amount of a scalable value (e.g. she is a bit upset, she is extremely upset), while the latter applies to ATTITUDE, ENGAGEMENT and to categories that are usually non-scalable from an experiential perspective; FOCUS grades meanings according to the prototypicality and preciseness by which the categorical boundary is drawn (e.g. she is a real friend, she is a friend to some extent). More comprehensive accounts of the GRADUATION resources at the level of lexicogrammar are provided by Eggins and Slade (1997: 133-137), Martin (1997), Martin and Rose (2007a: 42-48), and Martin and White (2005: 135-160). Hood and Martin (2007) further extend the GRADUATION network, specifying FORCE as embracing INTENSITY (of a quality), QUANTITY (of a thing), and ENHANCEMENT (of a process), and FOCUS encompassing VALEUR and FULFILLMENT. At the level of lexicogrammar, a great variety of linguistic resources including sub-modification (e.g. most important), processes infused with an attitudinal meaning (e.g. explore this area), and circumstances of manner (e.g. illustrate convincingly) can be deployed to intensify an attitude. QUANTITY may construe the meanings of ‘amount’ (e.g. considerable study) and ‘extent’ either as scope (e.g. for ten years, in the world) or distance (e.g. more recently, eastern vs. western societies) in relation to either time or space. The scaling of a process in terms of ENHANCEMENT can be achieved through a circumstance of manner (e.g. understand comprehensively). GRADUATION as FOCUS scales meanings along the dimensions of authenticity (e.g. a real killer), specificity (e.g. in particular), and fulfillment (e.g. fail to finish). GRADUATION is
recognized as an important resource in ‘flagging’ an attitudinal reading by scaling ideational meanings (Hood and Martin 2007). GRADUATION in multimodal texts will be discussed in relation to ENGAGEMENT values and attitudinal meanings in Chapter 4 and Chapter 5 respectively.

3.4 Systemic functional semiotic approach to visual communication

Over the past decade, Hallidayan social semiotic approach to language has been extended to include other semiotic resources such as visual images (Kress and van Leeuwen 1996, 2001, 2002, 2006; O’Toole 1990, 1994), sound (van Leeuwen 1999), spatial design (Martin and Stenglin 2007). As Kress and van Leeuwen state,

…we aimed at a common terminology for all semiotic modes, and stressed that, within a given social-cultural domain, the “same” meanings can often be expressed in different semiotic modes…we move towards a view of multimodality in which common semiotic principles operate in and across different modes. (Kress and van Leeuwen 2001: 1-2)

Inspired by the metafunctionally diversified perspective on language system, social semioticians have explored the mapping of multifunctionality across different modes of communication. This multifunctionality for the modalities of verbiage and image can be summarized in Table 3-4 below.

Table 3-4 Multifunctionality for the modalities of verbiage and image (Martin 2001: 311)

<table>
<thead>
<tr>
<th>metafunction:</th>
<th>naturalizing reality</th>
<th>enacting social relations</th>
<th>organizing text</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>modalities:</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>verbiage</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>-Halliday (1994)</td>
<td>ideational</td>
<td>interpersonal</td>
<td>textual</td>
</tr>
<tr>
<td><strong>image</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>-Kress and van Leeuwen (1996, 2006)</td>
<td>representation</td>
<td>interaction, modality</td>
<td>composition</td>
</tr>
<tr>
<td>-O’Toole (1994)</td>
<td>representational</td>
<td>modal</td>
<td>compositional</td>
</tr>
<tr>
<td>-Lemke (1998)</td>
<td>presentational</td>
<td>orientational</td>
<td>organisational</td>
</tr>
</tbody>
</table>
In the present section, I shall review the basic concepts of the descriptive framework proposed by Kress and van Leeuwen (2006), which will be adopted for visual analysis in the following three chapters. According to Kress and van Leeuwen (2006: 41-44), visual images, like all semiotic modes, fulfill the requirements for the representation of the experiential world, the communication between producers and receivers as well as the interaction between the participants represented in a visual design and its viewers, and the compositional arrangements of visual resources. As was outlined in Table 3-4 above, there are four major components in Kress and van Leeuwen’s model: representational resources for visually representing the material world, including the participants, events and objects involved in it; interactive resources that visually construe the interaction between viewer and what is viewed; modality judgements that deal with the reliability of visual messages; and compositional arrangements that are concerned with information value and visual emphasis. Among the four components, interaction and modality are concerned with the construal of social reality. Some relevant concepts of the other two aspects of visual communication will also be reviewed here, in that one of the important advantages of metafunctional perspective lies in the fact that it enables us to discuss interpersonal meaning in relation to the ideational (representational in visual displays) and textual (compositional) choices in the discourse.

3.4.1 Representation

Unlike the time-based semiotic modes such as language in which syntax is a matter of sequencing, the space-based semiotic modes such as image have ‘syntactic patterns’ that are encoded in spatial relationships (Jewitt and Oyama 2001). Kress and van Leeuwen (2006: 45-113) identify two types of structure in terms of representation: narrative structure and conceptual structure. What makes the difference between these two representational structures is the way the participants in the image are related to each other, i.e. whether it is based on the ‘unfolding of actions and events, processes of change’, or based on their ‘generalized’, ‘stable and timeless essence’
Theoretical Foundations

(Kress and van Leeuwen 2006: 59). The presence of a vector, an oblique or diagonal line that connects participants, can be taken as the visual mark for narrative representation, while absence of it indicates conceptual representation.

As Kress and van Leeuwen (1996, 2006) reveal, images involve two kinds of participants, i.e. represented participants that refer to the people, places, and things depicted in images, and interactive participants that are the image producer and viewer who communicate with each other via images. According to Kress and van Leeuwen (2006: 59-70), there are four types of process within the category of narrative representation, each of which has its own represented participants and circumstantial elements. An action process can be transactional when both Actor (i.e. the active represented participants from which the vector emanates) and Goal (i.e. the passive represented participants at which the vector is directed) are present and connected by a vector. An action process can also be non-transactional when the vector that departs from the actor does not point at any other participant. When the vector is formed by an eyeline of a represented participant, the process is a reactional one. The two represented participants involved in a relational process are termed Reacter and Phenomenon. When the eyeline vector emanates from the reacter but does not point at another represented participant, the reactional process is non-transactional. Narrative structure also includes speech process and mental process, which can be recognized by the presence of ‘dialogue balloons’ and ‘thought bubbles’ respectively. In these processes, the oblique protrusion of dialogue balloons and thought bubbles functions as the vector, connecting the represented participants to their speech or inner mental process. Therefore, the content of speech or thought is mediated through a sayer (in the case of speech process) or a senser (in the case of mental process). Still another narrative structure, i.e. the conversion process, involves a change of state in the represented participant. In this case, the goal of one action process is at the same time the actor of another action process, and this participant is referred to as relay. Narrative representations also involve circumstantial elements such as setting, means, and accompaniment.

The images that do not contain vectors are discussed under the heading of
conceptual representation (Kress and van Leeuwen 2006). Unlike narrative structures that focus on dynamic ‘doing’ or ‘happening’, conceptual structures are concerned with static ‘belonging’, ‘being’ or ‘meaning’. Three types of process can be identified in terms of class, structure and symbolic meaning (Kress and van Leeuwen 2006: 79-109). A classificational process relates represented participants to each other in terms of a taxonomy, with a set of participants distributed symmetrically across the picture space to demonstrate that they have something in common and thus belong to the same class. These participants act as the subordinates of another participant, which acts as their superordinate. A classificational process can be an overt taxonomy or covert taxonomy, depending on whether the superordinate is present in the image. The second type of process within the conceptual category is analytical process, in which participants are related based on a part-whole structure. The two types of represented participants involved in an analytical process are carrier (i.e. the whole), and possessive attributes (i.e. the parts that constitute the whole). If the carrier as well as its possessive attributes can be easily identified in the visual image (especially when labels are attached to the possessive attributes), they constitute a structured analytical process. On the contrary, if the image shows us only the unordered set of possessive attributes but not the carrier itself, the analytical process is considered to be unstructured. The third category, i.e. symbolic process, defines the meaning or identity of a represented participant, and can be further divided into symbolic attributive and symbolic suggestive. There are two participants in symbolic attributive, and the meaning of the participant is established in their relation; whereas only one participant is involved in symbolic suggestive, and the symbolic meaning is derived within the participant itself.

3.4.2 Interaction

In visual communication, there are resources for creating and maintaining the interaction between the viewer and what is represented within the picture frame. In doing so, an image engages the viewer and implies the position he/she should take in
Theoretical Foundations

viewing the visual displays. Kress and van Leeuwen (2006: 129) identify three main factors in the visual realizations of interactive meanings: contact, social distance, and point of view. Each of them encompasses options to specify the subtle relations between the viewer and the represented participants, which can be summarized in Figure 3-9 below. ‘Point of view’ is categorized into the ‘subjective attitude’ to describe the possibilities of expressing the socially determined attitudes towards the represented participants from both horizontal and vertical angles.

![Figure 3-9 Interactive meaning in images (Kress and van Leeuwen 2006: 149)](image)

In the visual system of ‘contact’ outlined in the model above, distinction is made between those images describing people or anthropomorphized things that look straight at the viewer and those images in which represented participants do not gaze at the viewer. Based on Halliday’s (1994) work on speech functions, Kress and van Leeuwen (2006: 117-118) use the term ‘demand’ pictures to refer to the images where an imaginary contact is established. They argue that the represented participants’ gaze symbolically demands that the viewer ‘enter some kind of imaginary relation’ with them, and the facial expressions and gestures of the represented participants further specify what is being demanded from the viewer. As for those images depicting represented participants as objects of the viewer’s detached observation, Kress and van Leeuwen (2006: 119-120) name them as ‘offer’ pictures. They
comment that no contact is made in this type of images, and thus the anthropomorphized creatures or inanimate objects are offered as ‘items of information’ to be contemplated by the viewer. The choice between ‘demand’ and ‘offer’, as is pointed out by Kress and van Leeuwen (2006: 120), is one way of suggesting the different relations between the viewer and the represented, so as to invite the viewer to engage with some and to keep them remain detached from others.

The basic opposition in terms of ‘contact’ is further elaborated and developed by Painter (2007) into a system network of visual focalization (see Figure 3-10 below), which the current research will draw upon. In this network, ‘contact’ and ‘observe’ are the technical terms for describing whether there is eye contact or not, which indicates how the viewer is positioned to engage with the represented participants (i.e. ‘character’ in Figure 3-10), or merely to observe what happens within the picture frame. Furthermore, the viewer of an image can be placed to view the visual world directly or vicariously through the eyes of a represented participant. Character focalization may occur across frame on adjacent pages or within a single frame. Within a picture frame, vicarious ‘observe’ can be achieved through viewing the world in the picture as what the represented participant himself/herself sees, or through viewing the visual world as well as the ‘focalizing character’ within the frame. The ways in which and eye contact is made will be considered in Section 4.4.1 of Chapter 4, while how character focalization is managed will be picked up in Section 5.3.2.2 of Chapter 5.

Figure 3-10 Visual focalization (Painter 2007: 47)
The second dimension of visual interaction is social distance, which is concerned with how images depict the represented participants as close to or far away from the viewer. This in turn is interpreted through the ‘size of frame’ in photographic technology. For example, the extreme close-up shot that shows only the face or head indicates intimate distance, and the close-up shot that include the head and the shoulders implies close personal distance. The medium close shot that cuts off the participants approximately at the waist suggests far personal distance, while the medium long shot that represents the whole figure expresses close social distance. Finally, the shots longer than that show the whole figure with space around it, or the torso of several people, which respectively correspond to far social distance and public distance (Hall 1964, 1966; see also Kress and van Leeuwen 2006: 124-125). The selection of a certain kind of social distance is determined by social relation, and the continuum of social distance also reflects the different social relations ranging from intimate, social, to impersonal. In other words, the participants can be represented as close friends, or as impersonal strangers that are more like general types than individuals with personalities. The association between putative reading position and social distance will be discussed in Section 5.3.2.1 of Chapter 5.

Turning to attitude in visual interaction, the present research primarily focuses on subjective, perspectival images, and skips the technical, objective drawings that ‘disregard the viewer’ (Kress and van Leeuwen 2006: 131). The resource of point of view enables the creation of the symbolic relations of involvement and detachment along the horizontal axis, as well as different power relations along the vertical axis. To be specific, the horizontal angle of frontality allows the viewer to be directly confronted with, and thus maximally involved with what is represented; whereas those images with the oblique angle depict the participants from the side as ‘others’ or ‘strangers’, positioning the viewer as a detached onlooker. In the case of the vertical angle, symbolic power is given to the viewer if low angle is taken, and hence the viewer looks down on the represented participant. If the represented participants are depicted from a high angle from which the viewer looks up at them, the symbolic
power is with the represented participants. Still another possibility is the eye-level angle, and in this case there is a relation of symbolic equality. As is explained by Jewitt and Oyama (2001), ‘power’, ‘detachment’ and ‘involvement’ are used for capturing the meaning potential of angles, and need to be activated by interactive participants; and secondly, symbolic relations are not real relations - they symbolically relate the represented participants to the interactive participants in the way that has been designed by the image-producer. The issue of point of view will be addressed as the analysis unfolds in the subsequent three chapters.

3.4.3 Modality

The concept of modality deals with one of the crucial issues in verbal and non-verbal communication, i.e. the reliability or credibility of messages. In everyday visual communication, photographs are generally believed to be the images that faithfully demonstrate things and people as they can be seen with the naked eye in reality. The truth value of photographs is based on the dominant, conventional, and most widely accepted definition of reality in contemporary times, which is referred to by Kress and van Leeuwen (2006: 165-166) as naturalistic coding orientation. To put it more technically, this reality standard is defined by the modern 35mm photographic technology, and any pictorial expression that exceeds or does not meet this standard is thought to be of low modality as far as naturalistic coding orientation is concerned. However, scientific graphs and diagrams that are abstract, general, and often colourless also claim to be real in technical context, owing to the fact that they reveal the essential qualities and look below the surface of things. They possess high modality value when viewed from another standard of reality, which is termed as abstract coding orientation according to Kress and van Leeuwen (2006: 165).

Coding orientations other than the aforementioned two standards can be identified in other communicative contexts. If an image displays a higher degree of colour saturation, a sharper contrast in terms of the play of light and shade, or a more detailed representation than the normal resolution of the 35mm colour photography, it may seem exaggerated, ‘more than real’ to a general viewing audience who hold
naturalistic coding orientation as their standard of truth. This type of image is often found in advertising, surrealist art, and horror films, in which the visual truth is based on the ‘pleasure principle’. What matters in this type of image is not the faithful appearance or the hidden essence, but the exciting, stirring, or soothing effects that the image designer aims to achieve. Kress and van Leeuwen (2006: 165) refer to this kind of visual truth as involving a sensory coding orientation.

Finally, when using an assembly instruction or a route map, people attach great importance to the practical use of the visual image as a blueprint. Here the usefulness or effectiveness of the image plays the essential role in this communicative contexts, and this truth criterion is technological coding orientation (Kress and van Leeuwen 2006: 165).

When analyzing the different degrees of articulation in different aspects of a given coding orientation, Kress and van Leeuwen (2006: 160-163) identify eight modality markers for describing the pictorial expressions in various aspects that are concerned with modality judgement: colour saturation, colour differentiation, colour modulation, contextualization, representation, depth, illumination, and brightness. Chapter 6 in the present thesis will focus on visual modality, with a multilevel view on the choice of different coding orientations in the primary and secondary EFL textbooks.

3.4.4 Composition

While representation deals with the way images represent the relations between participants within the picture frame and interaction is concerned with the relations between images and their viewer, composition relates the representational and interactive meanings into a meaningful whole through three interrelated systems: information value, salience, and framing (Kress and van Leeuwen 2006: 177).

Information value is realized through the placement of visual elements. Specifically, it is argued that different ‘zones’ of an image are endowed with different information values. For instance, the ‘left-right’ placement produces the ‘given-new’ structure since the conventional reading direction of a text in the modern society is
from left to right, from top to bottom\textsuperscript{10}. The element located on the left is regarded as the ‘given’, familiar, and agreed-upon point of departure for the message, while the element on the right is considered to be the ‘new’, problematic and contestable information that is not yet known and thus requires special attention from the viewer. Along the vertical axis from top to bottom within the picture space, what is placed in the upper part is presented as the ‘ideal’, generalized essence of the information that implies the promise of ‘what it might be’, whereas what is placed in the lower part is the ‘real’, practically-oriented, down-to-earth information that indicates the actual situation of ‘what it is’. In addition to the polarized compositions of ‘left-right’ and ‘ideal-new’, there is also centred composition where the opposition is between centre and margin. The central element provides the nucleus of the information, while the marginal elements are subservient to and organized around it.

To this point I have considered how the placement of elements in visual composition influences the information values of these elements. Salience, the second dimension in composition, deals with how some elements can be made more eye-catching, more conspicuous than others. Kress and van Leeuwen (2006: 202) identify several factors that can made an element salient, including size, sharpness of focus, tonal contrast, colour contrast, foregrounding and some other specific cultural factors. The third key aspect in visual composition, framing, is concerned with the disconnection and connection of visual elements. Disconnection of elements can be created through a wide range of devices, such as framelines, pictorial framing devices, empty space between elements, and discontinuity or contrast of colour. Connection of visual elements, on the contrary, can be achieved in various ways, including the absence of framing devices, the vectors that connect elements, and the similarities or continuities of colour.

Kress and van Leeuwen’s (2006) model provides an applicable descriptive framework for visual analysis. They further distinguish options and sub-categories within most of the concepts reviewed above. The brief introduction here aims at outlining the key principles and technical terms, while relevant illustrations and detailed explanations in relation to the data will be presented in Chapters 4 to 6 below.
3.5  Mapping interpersonal management in multimodal EFL textbook discourse

In the previous sections I have introduced the basic principles and key concepts of SFL, including the APPRAISAL system and the model for visual analysis informed by SFL. I am now in a better position to explain why APPRAISAL within SFL and the systemic functional semiotic approach to images provide robust tools for the exploration of interpersonal meaning in our data that involves both verbiage and image.

Following Halliday’s (1978: 2) social semiotic view on language as one of the semiotic resources that is interpreted within a sociocultural context, the current research views the EFL textbooks under discussion as a set of predictable configurations of semiotic choices from both verbal and visual meaning-making systems. As was mentioned in Section 3.2.4 above, studies of interpersonal meaning should take into consideration both the interactive and personal dimensions. In the remainder of this section, I shall explain theoretically how I model the interaction between textbook editor and reader, how to analyze the intermodal construal of attitudinal meanings, and how to develop a multilevel view of visual modality.

3.5.1  Identifying multimodal ENGAGEMENT devices and play of voices

As was mentioned in Section 3.4.1, there are represented participants (i.e. people and things represented within the picture frame) and interactive participants (i.e. image producer and viewer outside the picture frame) involved in an image. Therefore, three kinds of relations in visual communication: relations between represented participants, relations between interactive and represented participants, and relations between interactive participants (Kress and van Leeuwen 2006: 114). In the previous studies the first type of relation has been systematically approached under the heading of ‘representation’ (see Section 3.4.1 for further discussion), and the second type of relation has been comprehensively studied within the category of
‘interaction’ (see Section 3.4.2 above). However, the third type of relation in visual communication, i.e. how the image producer and receiver act on each other through images, seems relatively untouched.

It seems both natural and necessary for us to bring in the notion of ENGAGEMENT within the APPRAISAL theory (Martin and White 2005, White 2003) to fill the gap, when it is remembered that the system of ENGAGEMENT draws on Bakhtin’ concept of heteroglossia, which holds that every verbal communication is ‘dialogic’ and thus to speak or write is always to acknowledge what has been said/written before, and simultaneously to anticipate the intended reader/listeners’ responses (Martin and White 2005: 92-93). If this view is extended to multimodal communication that involves both language and images, the reader of verbal texts will at the same time be the viewer of visual images, whereas the producers of the multimodal discourse will include both writers and image producers. Considering the fact that all those resources which allow authorial voice to position itself with respect to other voices within a given communicative context are considered to be devices that encode ENGAGEMENT meanings, I shall attempt to identify multimodal ENGAGEMENT devices and analyze the ways they introduce alternative positions to construct and mediate the heteroglossic space.

Within EFL textbook discourse there is one type of voice that deserves our attention, i.e. the voice of the characters. Chapter 4 will be devoted to the analysis and discussion of the interplay of multiple voices by drawing on ENGAGEMENT, GRADUATION systems and theory on visual interaction, where I shall outline the ways that the represented participants (i.e. characters) as well as the interactive participants (i.e. both textbook editors and the intended reader) engage with one another.

3.5.2 Examining verbiage-image relations in attitudinal construal

As was reviewed in Chapter 2, a concern with ‘emotion and attitude’ has been highlighted in the new curriculum standards. However, to date few attempts have been made to consider how the multimodal meaning-making resources in textbooks can be deployed to realize the ‘emotion and attitude’ goal; this has been recognized as
a gap in the previous studies on textbooks in China (Zhang 2005). Although the significance of moral education and values education has been reflected upon by educationists (e.g. Zhong 2006; Zhu 2006), the discussions are seldom anchored onto the discourse analysis of pedagogic materials.

As was introduced in Section 3.3.1 above, the system of ATTITUDE within APPRAISAL provides discourse analysts with an analytical framework for the description of verbal attitudinal meanings. Recent advances on the system of ATTITUDE have gone beyond language to include other semiotic resources including visual images (e.g. Economou 2006; Tian 2007; White 2006). The visual texts that have been considered include lead pictures in newspapers, children’s picture books, and journalistic non-realistic visual displays. Chapter 5 of the present thesis will endeavour to enrich the research on ATTITUDE in multimodal texts by examining the intermodal attitudinal inscriptions and invocations in EFL textbook discourse, covering a range of visual styles from cartoon to photograph and portrait.

Taking into account both verbal and visual realizations of the three attitudinal regions (i.e. AFFECT, JUDGEMENT and APPRECIATION), I shall pay particular attention to the verbiage-image relations in encoding the evaluative stance. Based on the notion that JUDGEMENT and APPRECIATION are institutionalized or recontextualized AFFECT (Martin 2000a; Martin and White 2005), Chapter 5 will be devoted to the accumulative effect of the intermodal construal of the ‘emotion and attitude’ goal in the current pedagogic context. Gradability as a central property of attitudinal meanings will also be taken up in Chapter 5.

3.5.3 Developing a multilevel view on visual modality
In addition to the evaluative positioning that the putative reader is assumed to take, another issue closely related to the various styles of visual images in EFL textbook discourse is how true the message producer wants a representation to be taken by the receiver. I attempt to explore this issue by drawing on the concept of modality in social semiotics (Hodge and Kress 1988; van Leeuwen 2005), adopting modality markers and coding orientations (Kress and van Leeuwen 2006) as analytical tools.
By developing a multilevel view on modality in multimodal texts from primary, junior secondary, and senior secondary EFL textbooks, I aim at examining the ‘interdependence’ (Hodge and Kress 1988: 161) between modality and social relations.

As was made clear in Section 3.2 above, in SFL text and context are closely related to each other. This has greatly enhanced the explanatory power of SFL since it accounts for language use in relation to its social context, which can be further specified as variables of field, tenor, and mode. All semiotic modes simultaneously encode three types of meanings, i.e. ideational, interpersonal, and textual (Kress and van Leeuwen 2006: 41-44; see also Section 3.4 above). If the association between language patterns and social factors is projected onto other semiotic modes of communication, SFL may provide an equally effective tool to accommodate to the dual focus on both the semiotic resources and the social context. The influence of social relations (i.e. tenor) on the criteria for modality judgements will be investigated through a comparative analysis of three teaching units on a similar topic but for different educational levels. Furthermore, I shall attempt to discuss whether and in what way the visual arrangement in textbooks may be improved in relation to some trends in educational theory of the twentieth century.

Tenor is not the only factor affecting the choice of criterion for truth. Various degrees of deviation from the accepted truth criteria can be found in the multimodal texts targeting the same group of readers. Another register variable, i.e. field, will also be considered in Chapter 6 to discuss different choices of visual styles within a teaching unit as macrogenre.

3.6 Summary

This chapter is intended to provide the theoretical underpinnings for the present study. Some key tenets of SFL that are of high pertinence to this study have been reviewed. Those concepts include various assumptions – that language is one type of social semiotic and should be analyzed in relation to its social context; that the relation
between language and context is the probabilistic realization relationship; and that the system of language is structured according to its functions. Starting from these basic principles, I have concentrated the discussion on interpersonal semantics, with particular emphasis placed on the APPRAISAL system due to its relevance to the proposed analyses. The APPRAISAL theory is drawn upon to provide insights into both the negotiation of heteroglossic space and the construal of attitudinal meanings within EFL textbook discourse. New developments in SFL have taken into account semiotic resources other than language. It is held that all semiotic modes serve communicational as well as representational requirements and are structured according to specific compositional principles. This provides an analytical tool for dealing with our data, which involves images along with language.

With the purpose of articulating more explicitly the link between the aforementioned theoretical claims and the current research, I have also outlined the rationale of an SFL approach to the interpersonal semantics of multimodal EFL textbook discourse. This study aims to embrace both interactive and personal dimensions of interpersonal management, which can be further divided into three sub-studies covering the interaction of voices, attitudinal meanings, and modality. The following three chapters will be devoted to the detailed analyses and discussions of those aspects of interpersonal semantics, as is reflected in the multimodal EFL textbook discourse.

NOTES

1 The discussion here does not distinguish between ‘supervenience’ and ‘embedding’ (c.f. Martin 2008c).

2 As is pointed out by Matthiessen (1993), there are at least three ways of interpreting the relationship between context of culture and context of situation, which respectively follow three different organizing principles, i.e. rank, stratification, and potentiality. In addition to the multi-stratal model adopted in the present study, other models of context include the
single-stratum one that regards context of situation as an instance of the cultural potential (Halliday 1978, 1999), and the macro-micro relationship in which a culture can be viewed as consisting of situation types.

3 In SFL, the term register is initially defined by Halliday (1978: 111) as ‘the configuration of semantic resources that the member of a culture typically associates with a situation type. It is the meaning potential that is accessible in a given social context.’ Therefore, it is deemed as located at the semantic level of the linguistic system, whereas the categories of field, mode and tenor are regarded as the features of the context of situation and thus belong to one level up (Halliday in Thibault 1987: 610). The contextual model (Martin 1985, 1992, 1999) adopted here extends the notion of register to refer to the semiotic system that is metafunctionally organized and independently located between language and genre as one of the contextual strata.

4 Martin (1999) uses the stepped box diagram to distinguish the realization relationship between register and language from realization across strata within language.

5 In addition to realization and instantiation, another hierarchy that plays the complementary role in the SFL framework when exploring semantic variations is individuation (e.g. Martin 2008b). Individuation deals with the relationship between the reservoir of meanings in a given culture and the repertoire deployed by individuals in texts. The three hierarchies complement each other in that all systems along the realization hierarchy instantiate, and all individuate as well (Martin 2008b). To avoid too many foci, I primarily concentrate the discussion on the hierarchies of realization and instantiation.

6 According to Martin (2008b), coupling in its narrow sense involves those meanings that enter directly into semantic relations with one another, while broadly speaking it might refer to the meanings that co-occur more frequently than expected in a text or set of texts with reference to specifiable norms. Here I concentrate the discussion on the narrow reading of the concept of coupling.

7 According to Halliday (1992), there are two kinds of categories in the analysis of language: theoretical and descriptive. Those categories like ‘metafunction’, ‘system’, and ‘realization’ belong to the former, while those such as ‘material process’, ‘theme’, and ‘preposition’ which are language-specific and have evolved in the description of particular languages belong to the latter.

8 The semantic system that is currently and widely referred to as GRADUATION was termed AMPLIFICATION in some early works (e.g. Eggins and Slade 1997, Martin 1997).

9 The technical term ‘attitude’ here is different from the homonym within the APPRAISAL system. Here the ‘attitude’ refers to the subjective image and objective image in Western cultures since the Renaissance. The former has a ‘built-in’ perspective and its viewer can see what there is to see only from a particular point of view; whereas the latter does not have such a ‘built-in’ perspective and it reveals to its viewer everything there is to know even at the expense of the laws of naturalistic depiction. Most of the images in our everyday life are subjective, perspectival images, while objective images are acceptable in contexts such as assembly instruction and other
Theoretical Foundations

Based on the observation of our data, it can be inferred that the reading direction in most of the images under discussion follows the conventional reading direction in the contemporary society of China, i.e. from left to right rather than from right to left as was in the ancient times (cf. Jewitt and Oyama 2001).
4.1 Introduction

As was indicated in the previous chapter, interpersonal management includes both interactive and personal aspects. I shall start the exploration of interpersonal meaning in multimodal EFL textbook discourse with the interactive dimension. I explained in Section 3.5.1 the necessity and feasibility to draw on the ENGAGEMENT system to analyze the editor-reader/image producer-viewer alignment. This chapter will focus on the recognition of multimodal ENGAGEMENT devices and the analysis of the ways they contribute to the expansion and contraction of heteroglossic space, namely, the room for mediating the multiple voices in a text.

Specifically I shall firstly identify the multiple voices, i.e. editor voice, reader voice, and character voice, in multimodal EFL textbook discourse after a brief review of the studies on multiple voicing in texts. Then I shall analyze five types of multimodal resources as ENGAGEMENT devices in relation to the ways they function to mediate the play of multiple voices, which is proceeded by a discussion of visual interaction in terms of contact/observe, social distance and point of view.

4.2 Identifying multiple voices in multimodal EFL textbook discourse

The notion ‘voice’ that is explored in the current research needs to be clarified to
avoid possible confusion with its grammatical usage. The present study follows the fundamental and influential works by Bakhtin (1981, 1986) on heteroglossia – the dialogic and multi-voiced nature of all verbal communication, and draws on the semantic and rhetoric orientation that develops the ENGAGEMENT network of resources for the management of heteroglossic space in a text (Fuller 1998; Martin 1997, 2000a; Martin and White 2005; White 2003). Before the identification and analysis of the dialogic interaction between multiple voices in multimodal EFL textbook discourse, this section will begin with a brief review of the previous studies on voice, with a particular focus on the efforts made by systemic linguists.

**4.2.1 Previous studies on voice**

According to Bakhtin (1981: 281), all spoken and written texts are intrinsically multi-voiced and should be understood ‘against the background of other concrete utterances on the same theme’. As Bakhtin points out,

> Each utterance is filled with echoes and reverberations of other utterances. …

> Each utterance refutes, affirms, supplements, and relies on the others, presupposes them to be known, and somehow takes them into account. … each utterance is filled with various kinds of responsive reactions to other utterances of the given sphere of speech communication. (Bakhtin 1986: 91)

The notion of heteroglossia was later developed by Kristeva (1986) into the concept of ‘intertextuality’. In view of the focus of the present study, I shall mainly review the contributions that systemic linguists/semioticians bring to the studies of voice. Within SFL there are various perspectives on the issue of ‘voice’, which can be summarized as three main trends. Starting with the semiotic system of language, systemicists explore the heteroglossic nature of texts by analyzing the resources of projection at the level of lexico-grammar (Halliday 1994; Halliday and Matthiessen 2004; Martin and Rose 2007a; Thompson 1996). The projection resources that introduce other voices into a text include the verbal and mental processes that project
locutions and ideas (e.g. I think…, He doubt …), and the embedded projections realized in nominal groups (e.g. assertion, belief). Along with projection, other linguistic resources such as those of modality (e.g. probably, possibly) and concession (e.g. although) also bring in other voices and enable the negotiation between voices (Martin and Rose 2007a).

At the level of discourse semantics, a number of voice studies within systemic functional linguistics (henceforth SFL) view voice in its abstract sense, modelling it as the recurrent configuration of evaluative resources, or a collection of recurrent ‘syndromes of evaluation’ (Martin 2003) in texts. This notion of voice has attracted substantial attention within SFL. Iedema et al. (1994) study multiple voicing in media discourse, identifying ‘Reporter Voice’, ‘Correspondent Voice’ and ‘Commentator Voice’ in different types of media genres. Coffin’s (2000) work reveals the voice options of recorder, interpreter and adjudicator in history discourse. Macken-Horarik (2003) considers the writer/reader relations in students’ responses to narratives in examination situations, and points out that the attributes of successful responses include a sensitivity to the hierarchy of voices and values. Hood (2004) examines ‘voice’ in academic writing from two perspectives: the actual source of a given evaluation, as well as the abstract evaluative syndromes technically named as ‘voice roles’ (i.e. ‘Observer’, ‘Investigator’ and ‘Critic’). It is noteworthy that voice in the abstract sense is distinct from the concept of ‘voice’ in heteroglossia studies. The former deals with the sub-potential of evaluative meanings that are characteristic of a specific register (e.g. ‘reporter voice’ in media discourse), which can be explicated as ‘key’ in the cline of instantiation concerning evaluation (Martin and White 2005: 164); whereas the latter is concerned with the sources of propositions and evaluations that are heteroglossically present in texts (e.g. the ‘voice’ in Bakhtin’s notion of dialogism).
Table 4-1 Cline of instantiation – evaluation (Martin and White 2005: 164)

1. APPRAISAL (system) – the global potential of the language for making evaluative meanings, e.g. for activating positive/negative viewpoints
2. key (register) – situational variants or sub-selections of the global evaluative meaning making potential – typically reconfiguration of the probabilities for the occurrence of particular evaluative meaning-making options or for the co-occurrence of options
3. stance (text type) – sub-selections of evaluative options within text; patterns of use of evaluative options within a given ‘key’ associated with particular rhetorical objectives and the construction of authorial personae
4. evaluation (instance) – instantiation of evaluative options in text
5. reaction (reading) – the take-up of evaluative meanings in a text according to the listener/reader’s subjectively determined reading position; the attitudinal positions activated by the reader as a result of their interaction with the text

The perspective on voice explored in the current research centres around voice in the sense of source, that is, whether a given proposition or evaluation is attributed to the author or to another source that is co-present in the text. In particular, the present examination owes most to the taxonomy of ENGAGEMENT meanings (known as the semantic system of ENGAGEMENT) and the ways in which engagement meanings can be graded (known as GRADUATION) developed as part of APPRAISAL theory within SFL (Martin 2000a; Martin and White 2005). Informed by Bakhtin’s notion of heteroglossia, the ENGAGEMENT system (Martin and White 2005; see also White 2003) sets up networks of options for opening up or closing down heteroglossic space for multiple voices in a text. Given the large number of studies that provide semantic and functional account of the linguistic resources that encode intersubjective positioning, to date, few studies can be found on examining the ways in which multimodal meaning-making resources function to expand or contract the heteroglossic space. As was mentioned in Chapters 1 and 2, modern communicative channels are far from being monomodal. The current research strives to examine the role of multimodal ENGAGEMENT devices in mediating the play of multiple voices in the textbook discourse under the current EFL teaching context in China. Before embarking on the detailed analysis and discussion, I shall first identify the multiple
voices that exist in the multimodal EFL textbook discourse.

4.2.2 Multi-voicing in EFL textbook discourse

In this section three voices that constitute the heteroglossic backdrop of EFL textbook discourse will be identified, i.e. editor voice, reader voice, and character voice. Editor voice can be easily observed in the use of imperative clauses as section titles, which has found wide application in the EFL textbooks under discussion. These imperative clauses are either the unmarked ones that have no Mood such as *Read and write*, *Write and say*, *Match and say*, and *Write and draw*, or the *let’s* form where the understood Subject is ‘you and I’ (see Halliday 1994: 87, Thompson 1996: 49) like *Let’s start*, *Let’s learn*, *Let’s talk*, *Let’s read*, *Let’s play*, *Let’s do*, *Let’s make*, *Let’s check*, *Let’s find out*, *Let’s draw*, *Let’s sing*, and *Let’s chant*. As Martin and White (2005: 110-111) point out, in contrast with ‘deontic’ modality (e.g. You *must* read the following paragraph.) that explicitly treats the demand as an assessment by the speaker of obligation instead of as a command, imperative is monoglossic in the sense that it does not refer to or allows for other possible alternative actions. On the other hand, if treated like modulation, imperative opens up monoglossic setting to include the reader, and hence the speaker’s (i.e. the editor’s) role as a participant in the dialogic exchange is recognized. This type of section titles encodes editor voice by way of giving general directions on what the reader is supposed to perform in certain tasks.

In addition to imperative clauses, the section titles in the EFL textbooks can also be nominal groups such as *Pair work*, *Group work*, *Story time*, *Pronunciation*, *Task time*, and *Culture*. This type of section titles encodes editor voice by clarifying the purposes or goals of teaching in the sections. Meanwhile, it implies the actions that the reader is assumed to take in the teaching section (e.g. *Group work* implies ‘You *must* carry out the task in groups’). Through treating the proposal as an entity, this kind of section title ‘backgrounds’ the proposer and conceals the ‘must-ness’ of the proposal, which tends to bear some resemblance to ‘demodalization’ in administrative discourse (Iedema 1995). Demodalization can be achieved in various ways, ranging
from ‘nominalization’ (i.e. naming the commanding action) such as *Pronunciation*, ‘demodulation’ (i.e. masking the proposer and the controlling nature of the command) such as *Pair work* and *Story time*, to ‘generalization’ (i.e. generalizing the demodulations as members of a higher category) like *Culture*. Through ideationalizing the proposal, the process of control and demand is disguised as a natural and non-negotiable rule or institutional entity, and thus the institutional power of the editor is further enhanced.

Section titles are by no means the sole place where editor voice can be detected. The practice of attaching labels to certain objects in a visual image, which is termed as ‘labelling’ in the current study, can equally convey editor voice. Take Figure 4-1 for instance. It is adapted from an image in *PEP Primary English Students’ Book I for Year 4* (2003: 9). In this image, the unfinished words _ce-cream, _ish, _oose and _amburger are inserted as labels into an image that depicts a chef holding a tray full of dishes, guiding the reader to complete the words and the drawings. Labelling is recognized as a kind of ENGAGEMENT device in multimodal EFL textbook discourse. I shall probe into this issue in Section 4.3.1 later in the present chapter. Nonetheless, the labelling here in Figure 4-1 is different from labelling in the general sense, in that it contains a missing letter that is intended to involve reader’s participation. This type of intentionally unfinished text is classified into the category of ‘jointly-constructed text’, which will be discussed in detail in Section 4.3.3 below.
One of the distinctive features of the series of primary EFL textbooks under discussion is the extensive use of cartoon characters in creating the context of language in use. These characters either demonstrate how to play certain games that are considered to be conducive to language learning, or guide the reader in performing certain tasks. At the beginning of each primary EFL textbook there is a multimodal page introducing the major characters in the textbook (see for example Appendix I), presenting the names and images of fictional students and teachers (e.g. Chen Jie, Mike, and Miss White) as well as the drawings of anthropomorphized animals (e.g. the bear Zoom and the fox Zip).

Character voice is often conveyed through a speech process (Kress and van Leeuwen 2006), which is visually realized by a dialogue balloon with an oblique line linking the sayer (e.g. the character) to the content of the speech. Kress and van Leeuwen (2006: 68) describe this kind of structure as ‘projective’, as the utterance is not represented directly, but mediated through a sayer. Take Figure 4-1 again for example. The dialogue balloon or speech balloon, with its oblique protruding line, connects the character (i.e. Mike in the guise of a chef) with the utterance (i.e. *What are they? Do you know?*). Dialogue balloon is identified as one type of ENGAGEMENT device, which will be further analyzed and discussed in Section 4.3.2 in this chapter.

I now turn to the third voice in EFL textbook discourse, i.e. reader voice. As Martin and White (2005: 162-164) point out, the subjectively determined reading position is the end point of instantiation, which activates the attitudinal positions in a text. As is observed in the data, the reader of an EFL textbook is not merely passive addressee. Reader voice is explicitly articulated in jointly-constructed text, which involves or requires reader’s participation in its ultimate completion. Take Figure 4-1 for instance. The drawings of empty dishes and the blanks in the unfinished words _ce-cream, _ish, _oose and _amburger are left for reader’s completion. In SFL, language is regarded as a semiotic system of meaning potential, and language behaviour is interpreted as choice (Halliday 1978: 39; see also Eggins 2004: 20). Nevertheless, language is only ‘one of the semiotic systems that constitute a culture’
(Halliday 1978: 2). If this principle of potentiality is extended to include other semiotic systems, it can be argued that in making a choice from a semiotic system to finish a jointly-constructed text, what the reader actually writes or draws (i.e. the actual semiotic choice) gets its meaning by being interpreted against the background of what he/she could have written or drawn (i.e. potential semiotic choices). Jointly-constructed text, which is essential in engaging the reader in multimodal EFL textbook discourse, will be discussed in detail with examples in Section 4.3.3 below.

There are two additional resources for managing heteroglossic space in multimodal discourse, i.e. illustration and highlighting. The subsequent section will be devoted to the examination of the ways in which these five types of multimodal resources realize the ENGAGEMENT meanings of [disclaim], [proclaim], [entertain] and [attribute].

4.3 Multimodal resources as ENGAGEMENT devices

In the present section I shall attempt to identify and analyze the multimodal ENGAGEMENT devices for managing heteroglossic space in multi-voiced EFL textbook discourse, with a discussion of the ENGAGEMENT meanings, i.e. [disclaim], [proclaim], [entertain] and [attribute] that these multimodal ENGAGEMENT devices encode.

4.3.1 Labelling

As was mentioned in Section 4.2.2 above, the term ‘labelling’ used in the present study refers to the practice of inserting certain labels into a visual image to indicate some information concerning certain objects in the image. In primary and junior secondary EFL textbooks, labelling can be frequently observed in those images that depict the scene of a story which may provide a context for language in use. See Examples (1) and (2) below for instance.
Both Examples (1) and (2) embrace images of ‘narrative representation’ (Kress and van Leeuwen 2006: 45). Example (1) is extracted from *Unit 3 Is this your skirt?* in *PEP Primary English Students’ Book II for Year 4* (2003: 31), depicting the scene in which the character *Amy* is looking for her white socks in haste in her bedroom. The image encompasses an action process, with *Amy* being the Actor and clothes the Goal, while *Amy*’s arms serve as the two vectors. Watching the messy bedroom, *Amy*’s mother is anxious to help. Thus the image also includes a reactional process, with *Amy*’s mother as the Reactor and the scene in the bedroom as the Phenomenon. The vector in this reactional process is formed by the eyeline of *Amy*’s mother. Speech process is also involved. The two characters’ voices are conveyed through dialogue.
balloons.

As is shown in the image, editor voice loses no time in negotiating meanings with character voice via labelling. By putting the labels of *jeans, pants, shorts, socks,* and *shoes* onto the clothes scattered around the room, editor voice challenges or rules out alternative positions and thus limits the range of choices. For example, the label *shoes* fends off such alternatives as *sneakers* or *sports shoes.* If put into words, the meaning of the visual practice of labelling can be presented in the clause ‘*I contend that they are called shoes.*’ Therefore, it may be inferred that labelling functions to contract the heteroglossic space, and can be identified as one of the resources that encode [proclaim], or [pronouncement] to be more precise. By overtly pronouncing the names of the items portrayed in the image, editor voice intervenes in the visual narrative representation. The multimodal resource of labelling can be viewed as ‘an interpolation of the authorial presence so as to assert or insist upon the value or warrantability of the proposition’ (Martin and White 2005: 128).

Example (2) is taken from *Unit 7 How do you make a banana milk shake?* in the junior secondary EFL textbook *Go for it Students’ Book I for Year 8* (2005: 42). In Example (2), through the use of labels that indicate the names of the ingredients, editor voice intrudes upon the image that depicts *Marie* and *Katie* making fruit salad. The reader of the textbook is required to fill in the table on the left of the image with the labels proclaimed by editor voice. The proclamations like *yogurt* and *apples* actually rule out other possibilities such as *milk* and *strawberries,* and such emphasis or insistence implies some degree of resistance to any potential opposing viewpoint. Therefore, the formulation of labelling recognizes the heteroglossic diversity in the communicative context through presenting editor voice as challenging against any dialogic alternative. These labels function to close down the heteroglossic space in the text, and help concentrate the reader’s attention on the pronounced vocabulary items, which are supposed to be part of the language goals that are prescribed to achieve.

As was analyzed above, the ENGAGEMENT resource of labelling by which editor voice is interpolated into multimodal discourse so as to stress the prescribed teaching
goals while confronting possible contrary positions. These instances of pronunciation, according to Martin and White (2005: 129), are dialogic in the sense that they acknowledge the presence of counter viewpoints in the given communicative setting, and at the same time they are also contractive due to their challenge or resistance to possible dialogic alternatives. It is worth mentioning that this labelling practice cannot be found in senior secondary EFL textbooks. In other words, editor voice chooses other means to negotiate meanings with character voice in the senior secondary English-teaching context, in which it is mainly the verbal texts, rather than the visual images, that provide the communicative context. I shall return to this issue in Section 4.3.5 below, where the ENGAGEMENT resource of highlighting will be examined.

### 4.3.2 Dialogue balloon

Dialogue balloon, with its oblique protruding line connecting the sayer with his/her utterance, are commonly found in the primary and junior secondary EFL textbooks. As was discussed in Section 4.2 above, this type of visual structure is referred to as projection due to the indirect representation of the utterance. Dialogue balloon introduces character voice into a text, thus encoding the ENGAGEMENT meaning of attribute. In EFL textbook discourse three types of dialogue balloon are identified based on the different functions they perform: lending support to editor voice, explaining rules of games by demonstration, and giving directions to the reader.

#### 4.3.2.1 Lending support to editor voice

In bringing the external character voice into a text, certain dialogue balloons function to lend support to editor voice in the verbal text. See Example (3) below for instance, which is a teaching section entitled ‘Good to know’ taken from Unit 2 My favourite season in PEP Primary English Students’ Book II for Year 5 (2003: 24).
The upper part of Example (3) is verbally describes seasonal differences between Beijing and Sydney. The visual image underneath the verbal text demonstrates a dialogue between the two characters, i.e. Chen Jie and John. John asks Chen Jie what season it is in March in Beijing. She says it is spring and asks him what season it is in Sydney. He then says it is fall. The dialogue balloons bring in the external character voice, associating the proposition advanced by editor voice within the verbal text (i.e. In Beijing, it’s spring from March to May. Summer is from June to August. Fall is September to November. Winter is December to February the next year. But, in Sydney, it’s spring from September to November. Summer is from December to February the next year. Fall is from March to May.) with the external source of support from the characters who are assumed to be two students from China and Australia. Consequently, the dialogue balloons contribute to the heteroglossic alliance in the text, and therefore realize the ENGAGEMENT meaning of [attribute], or [acknowledging] to be more specific. By introducing character voice that is in support of the argument, editor voice presents the proposition as highly credible, hence aligning the reader into the position held by the editor.

4.3.2.2 Explaining rules of games by demonstration

In addition to giving support to editor voice, some dialogue balloons can also function
to explain the rules of games by demonstration, as is illustrated in Examples (4) and (5).

**Example (4)**

![Image of a cartoon showing two children playing a game. The dialogue balloons read: Can I have a Coke, please? Here you are. Thank you.]

[PEP Primary English Students’ Book I for Year 3  2003: 53]

Example (4) is an excerpt from *Unit 5 Let’s eat* in the *PEP Primary English Students’ Book I for Year 3* (2003: 53). Through the use of cartoons, it shows the reader (i.e. a primary school student) how to practice the expressions of commanding and offering by playing a game with pictorial cards. In Example (4) there is no verbal instruction that tells the reader how to perform, but a visual demonstration is given through the use of two images that involve a dialogue between the characters Sarah and Wu Yifan. Their utterances are indicated in three dialogue balloons, and the two visual images as a whole vividly demonstrate how to play the game,

**Example (5)**

![Image of a survey page with questions about the healthiest student. The questions include: How often do you eat vegetables? What sport do you play? Maria exercises every day. She likes to play...]

[Go for it Students’ Book I for Year 8 2005: 5]
encouraging the reader to play a similar one. Therefore, it can be inferred that the dialogue balloons in the visual images function to introduce character voice, and thus contribute to the explanation of the rule of the game via a visual demonstration.

In Example (5) two visual images lie beneath a verbal text that serves as an instruction given by the editor to show how to conduct a mimic survey concerning *Who is the healthiest*. Through the use of three imperative clauses (i.e. *add five questions to the survey on page 81, then ask three classmates the questions and take notes, and discuss and decide*) and one WH-interrogative clause (i.e. *who is the healthiest student*), the editor demands action and information from the reader. The accompanying two visual images provide a visual demonstration to support the verbal text. The image on the left (henceforth Image I) takes a ‘public distance’ shot (Hall 1966: 110-120, cited in Kress and van Leeuwen 2006: 124), and includes within the picture frame the figures of eight students who are interviewing each other and taking notes. The other image on the right (henceforth Image II) portrays at the ‘close personal distance’ (ibid) a student reporting his findings to the class. While the verbal text primarily conveys editor voice, the two visual images mainly express character voice. A corresponding relationship can be found between the two semiotic resources. Take speech function (Halliday 1994: 69) for instance. Table 4-2 below summarizes the speech functions in the verbal text of Example (5) and their corresponding visual patterns in Images I and II.

<table>
<thead>
<tr>
<th>speech functions</th>
<th>verbal instances (editor voice)</th>
<th>corresponding visual patterns (character voice)</th>
</tr>
</thead>
<tbody>
<tr>
<td>command</td>
<td><em>add five questions</em></td>
<td>In Image I, a demonstration of two of the questions is given, as is presented in the dialogue balloons (i.e. <em>How often do you eat vegetables</em> and <em>What sports do you play</em>).</td>
</tr>
<tr>
<td>Command</td>
<td><em>ask three classmates the questions, take notes, discuss</em></td>
<td>In Image I, students ask each other questions, taking notes and having free discussion in groups of four people.</td>
</tr>
</tbody>
</table>
It can be inferred from Table 4-2 that the verbal instruction and visual demonstration in Example (5) work in tandem to encourage the reader to carry out a similar survey about habits and health, with the visual demonstration already setting a good example for them to follow. Here character voice echoes with editor voice in the way that the dialogue balloons demonstrate what kind of questions should be added to the survey and how a report should be made according to the findings of the survey. Therefore, it may be justified to say that the dialogue balloons in Example (5) have the dual functions of explaining the rules of games by demonstration as well as lending support to editor voice. The images as a whole, on the one hand, can be read as an elaboration (Halliday 1994) of the verbiage, that is, as an example of how to conduct the survey, while on the other, they work together with the verbal texts in giving instructions to the reader, though in an implicit way.

It is noteworthy that the visual demonstration with dialogue balloons can be frequently observed in the task-oriented teaching sections such as Let’s play, Task time, Pair work and Group work. In those teaching sections the interaction between readers is required in fulfilling the tasks, and at least one way of accomplishing the tasks is vividly demonstrated by the characters in the visual images. Here the dialogue balloons actively make allowance for character voice, hence opening up the heteroglossic space and encoding the ENGAGEMENT meaning of [attribute].

**4.3.2.3 Giving directions to reader**

In most images in the EFL textbooks under discussion, the represented participants (i.e. the characters), do not gaze out at the viewer (i.e. the reader of the textbook). As was mentioned in Section 3.4.2 in Chapter 3, these images belong to the ‘observe’ type (Martin 2008a; Painter 2007; cf. Kress and van Leeuwen 2006), where the reader
is positioned to simply observe what is represented within the picture frame. In our data there are a few images in which the characters look directly at the viewer and thus eye-to-eye engagement is created as far as focalization is concerned. As for this ‘contact’ type of images, directions are often given in the corresponding verbal texts through character voice, which is indicated in the imperative or interrogative clauses in the dialogue balloons. The following example is taken from *Unit 1 How do you go there* in *PEP Primary English Students’ Book I for Year 6* (2003: 12).

(6)

![Good to know](image)

In Example (6) the little policeman in the image looks at the viewer outside the picture frame, and therefore an eye contact is set up between the character and the viewer. The character’s gaze symbolically invites the viewer to engage in an imaginary relation. The language in the dialogue balloon (i.e. *Look, read and match*) gives directions to reader, clarifying what is required from the reader in completing the exercise. In the case of Example (6), the reader is required to match the traffic signs with their corresponding meanings.

Figure 4-1 in Section 4.2.2 above is another example that illustrates the function of dialogue balloons in giving directions to reader. In the image of Figure 4-1, the character *Mike* looks into the viewer’s eyes. Through the gesture of holding the tray full of dishes and the use of a WH- interrogative clause and a yes-no interrogative
clause in the dialogue balloon, the character encourages the viewer to draw the dishes and complete the words. Example (6) and Figure 4-1 belong to jointly-constructed texts, in which the reader’s contribution is essential in the completion of the texts. I shall discuss in detail jointly-constructed texts in the section that follows.

4.3.3 Jointly-constructed text
Another multimodal ENGAGEMENT resource that encodes heteroglossia in EFL textbook discourse is the jointly-constructed text. The term ‘jointly-constructed text’ used in the current research refers to any text that is intentionally unfinished and aims to involve reader’s participation in its ultimate completion. Jointly-constructed text, which is essential in aligning the reader, has found wide application in the primary, junior and senior secondary EFL textbooks under discussion. Jointly-constructed text takes a great diversity of forms, and the multimodal modes of communication further enrich the ways of engaging reader voice. In the present section I shall attempt to analyze and discuss the multimodal jointly-constructed text as an ENGAGEMENT device in the EFL textbooks for the three different educational levels.

4.3.3.1 Jointly-constructed drawing exercise
Example (7) is an excerpt from Unit 2 Look at me in PEP Primary English Students’ Book I for Year 3 (2003: 16). It is a multimodal exercise whose main body is an unfinished picture of a human face. The human face image is an analytical process (Kress and van Leeuwen 2006: 87) to be completed. As was reviewed in Section 3.4.1 in Chapter 3, there is a part-whole relation between the participants (i.e. carrier and possessive attributes) in an analytical process. In Example (7), the picture of human face is the carrier, while the eyes, ears, nose and mouth to be drawn by the reader are the possessive attributes. The labels eye, ear, nose and mouth proclaimed by editor voice indicate what should be drawn and where they should be drawn. The reader is thus required to follow the labels to complete this structured analytical process.
The strong, diagonal line of the pencil at the top left corner of the human face image forms a vector, indicating the presence of a narrative representation. According to Kress and van Leeuwen (2006: 59-63), the vector links Actor and Goal in a transactional visual narrative process, with the Actor being the participant from which the vector departs and Goal as the participant at which the vector is directed. In some circumstances, the Actor can be fused with a vector to different extents. In the case of Example (7), the salient pencil, which is placed in the foreground with full colour saturation, plays the dual role of both the Actor and vector. The Goal in this transactional action process is the unfinished drawing, and the vector formed by the oblique pencil encourages the reader to participate in co-constructing the multimodal text.

I shall further examine this jointly-constructed text in terms of visual composition. At the top right corner of the whole multimodal text there is another image in a smaller size depicting the character Mike holding a finished picture of human face. The upper right position that the smaller image occupies indicates that it is presented as the Ideal and the New (Kress and van Leeuwen 2006: 179-194). As compared with the upper smaller image, the larger unfinished human face is placed in lower left position, which suggests its status as the Real and the Given (ibid). In other words, the unfinished drawing is presented as the practical, agreed-upon starting point, whereas the finished picture is the idealized and generalized
information that is unknown to the viewer (i.e. the textbook reader). In light of the labels that indicate the possessive attributes and the contour of human face, the reader is supposed to achieve the distant, ideal goal as is indicated in the upper smaller image. In this multimodal text the practice of labelling conveys editor voice, whereas the upper smaller image expresses character voice (which will be analyzed as the ENGAGEMENT device of ‘illustration’ in Section 4.3.4 below). The lack of completion or fulfillment in this multimodal jointly-constructed text in effect opens up space for other viewpoints or voices (in this case, reader voice) to join in. Reader voice engages with editor voice and character voice through co-constructing the text. Owing to the fact that the answer to this multimodal exercise comes from the external voice of the reader, the meaning of the unfinished jointly-constructed text can be expressed in the clause ‘According to the reader, the picture will be …’. Therefore, it can be inferred that jointly-constructed text functions to encode the ENGAGEMENT meaning of [attribute], expanding the heteroglossic space by bringing in reader voice.

### 4.3.3.2 Multimodal jointly-constructed herald page

I now embark on the investigation into a type of commonly found multimodal jointly-constructed text in junior secondary EFL textbooks, which always appears at the very beginning of each teaching unit. Consequently, this type of jointly-constructed text is referred to as ‘multimodal jointly-constructed herald page’ in the current research. Example (8), which is taken from Unit 10 Where did you go on vacation in the junior secondary textbook Go for it Students’ Book II for Year 7 (2005: 59), provides a good example.
The herald page in Example (8) is part of SECTION A in the teaching unit, consisting of a visual image as well as three exercises (i.e. 1a, 1b and 1c). Editor voice is explicitly conveyed in the Language Goal on top of the image and in the instructions of the three exercises:

Language Goal: Talk about past events.
1a. Match the activities with the pictures [a-g].
1b. Listen and number the people [1-5] in the picture.
1c. PAIRWORK Where did the people in activity 1a go on vacation? Make conversations.

Four imperative clauses and one WH-interrogative clause are employed in the Language Goal and in the three instructions. The imperative clauses all take the
unmarked form without Mood. Here the power of the editor can be easily perceived through the speech functions of command and question in the clauses in the Language Goal and instructions. The analysis of the visual composition may bring better insight into the understanding of editor power in this multimodal text. When viewed along the vertical axis, the Language Goal is placed on top of the visual image, which is the visual position that implies the ideal status, indicating the aspiration and desire the reader is supposed to possess in accomplishing the goal. In other words, the upper position of the Language Goal further enhances the goal-setting and leading role of the editor.

The visual image underneath the Language Goal provides further specification and elaboration, and serves as a communicative setting for the newly-learnt expressions that are used for talking about past events. In this image seven characters are portrayed as sitting or lying on the grass, recollecting and talking about past experiences and activities. Character voice is brought into the multimodal text via two dialogue balloons and seven thought balloons. The utterances in the dialogue balloons verbally demonstrate how to ask and tell other people about past events. The function of a thought balloon is quite similar to that of a dialogue balloon, in that the bubbled vector protruding from a thought balloon links the inner mental process with the senser, and hence the content of thoughts is mediated through the senser instead of being represented directly. However, in Example (8), unlike the dialogue balloons that draw upon verbal utterances, the thought balloons visualize the past activities that the characters recall to mind. Moreover, a letter label is attached to each of the images within the thought balloons, which indicates editor voice’s intrusion into character voice. Meanwhile, these letter labels also function to invite the reader’s attention to the relation between the images and the corresponding verbal expressions in exercise 1a. In addition to the seven letter labels, five square boxes are also inserted and left blank for the reader to fill in according to the tape recording in exercise 1b. Through doing the matching exercises in 1a and 1b, reader voice actively participate in the final completion of this multimodal herald page.

Example (8), like any other multimodal herald page, is placed at the start of a
new teaching unit. At that beginning stage of studying a teaching unit, there is an unequal possession of knowledge between the editor and the reader. That may explain why a power relationship is enacted between the communicative parties in the goal-setting part and instructions, as was shown in the analysis above. However, along with the strong editor voice, solidarity is still established and enhanced through the interpolation of character voice and reader voice into the multimodal text. Through the uses of dialogue balloons and thought balloons, a variety of situations for practicing talking about past events are provided but mediated by the sayers and sensers. These sayers and sensers are the characters of the textbooks, represented in a comic style as the peers of the reader (i.e. a junior secondary students) so as to enhance intimacy and cordiality. In other words, the interpolation of character voice sets up the multimodal text as not merely pronounced by editor voice, and hence reduces the unfamiliarity of what will be taught in the teaching unit. Furthermore, the herald page is not yet completed until the reader’s participation is taken into account, which contributes to the expansion of the heteroglossic space. To sum up, in Example (8) the multimodal ENGAGEMENT devices of dialogue balloon, thought balloon and the jointly-constructed text work together to construe the communicative setting as heteroglossic rather than monoglossic with only one voice.

4.3.3.3 Jointly-constructed verbal text with accompanying images

In this section, I shall examine a third type of multimodal jointly-constructed text, in which the reader participates in constructing the verbal text in light of the accompanying images. Take Example (9) for instance, which is taken from Unit 4 Earthquakes in the senior secondary EFL textbook New Senior English for China Student’s Book 1 (2004: 30).
In Example (9) editor voice is explicitly conveyed in the instruction that lies above a speech draft as well as four stamps that depict the new Tangshan:

Imagine that after your speech, Zhang Sha asks you to give a short talk about the new stamps to honour the city. You may use the model or write your own little talk.

By using an imperative clause, editor voice directs the reader to envisage preparing for a short talk about the four new stamps designed to honour the Tangshan City. The use of a finite modal operator may in the second clause indicates the low degree of obligation. In other words, the reader is allowed to make use of the model provided in the draft, while the alternative choice of not using the draft is still valid. The character Mr Zhang Sha is verbally introduced into this text but not visually represented. His identity was indicated in the previous teaching sections as an officer from the government of Tangshan City. Through the introduction of this external character voice, the heteroglossic space of the text is expanded. Reader voice is invited to engage interactively with editor voice as well as character voice, in that the exophoric references of the second person you and the possessive deictic your associate the external reader voice with the proposal being advanced in the verbal instruction. Nevertheless, the jointly-constructed speech draft will not be finalized until the reader’s contribution is considered. Due to the focus of the present chapter,
I stop at the identification and description of the play of voices in Example (9) at this stage. I shall return to this jointly-constructed multimodal text in Section 5.4.3 in Chapter 5, where emphasis will be placed on text-image relations in terms of the multimodal construal of attitudinal orientation in jointly-constructed text.

To sum up, the multimodal resource of jointly-constructed text brings in reader voice, and it is identified in the ENGAGEMENT network as encoding [attribute]. Character voice is often introduced into the jointly-constructed text through a variety of semiotic modes, and the relationship of power and solidarity in EFL textbook discourse can be negotiated through the play of the multiple voices.

### 4.3.3.4 Readership construal

Before continuing to conduct the analyses of other types of multimodal ENGAGEMENT resources, I attempt to examine the construal of readership in jointly-constructed texts in the present section. As was reviewed in Section 3.2.3 in Chapter 3, it is the act of reader interpretation that is at the end of the ‘cline of instantiation’ from system to reading (Martin and White 2005: 163; see also Martin and Rose 2007a: 312). In other words, the lowest level of instantiation is the meaning taken from the text based on the social subjectivity of a reader, rather than the text itself. The cline from the system of meaning-making potential to the final reading position is summarized in Table 4-3.

Following Martin and White’s (2005) view on alignment and putative reader, I consider that the jointly-constructed texts construe for themselves an ‘ideal’, ‘intended’ or ‘envisaged’ reader[^3], with whom the textbook editor aims to align. As was analyzed above in the previous section, jointly-constructed texts open up the space for the dialogic alternative from the reader, thus aligning the reader in the final completion of the texts.
Table 4-3 Cline of instantiation – from system to reading (Martin and White 2005: 163)

<table>
<thead>
<tr>
<th>Cline</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. system</td>
<td>(the global meaning making potential provided by the language)</td>
</tr>
<tr>
<td>2. register</td>
<td>(contextual variants or sub-selections of the global meaning making potential – involving more fully institutionalised reconfigurations of the probabilities for the occurrence of particular meaning-making options or for the co-occurrence of options)</td>
</tr>
<tr>
<td>3. text type</td>
<td>(groups of texts with comparable configurations of the probabilities of occurrence of options - involving less fully institutionalised configurations of the probabilities)</td>
</tr>
<tr>
<td>4. instance</td>
<td>(individual texts – the actualisation of the global meaning making potential, typically in conformity with the sub-potential settings of a given register)</td>
</tr>
<tr>
<td>5. reading</td>
<td>(the uptake of meanings in a text according to the listener/reader’s subjectively determined reading position.)</td>
</tr>
</tbody>
</table>

A text construes or models its reader by presenting the writer as taking for granted that the reader shares with the writer a particular viewpoint, or as assuming that the proposition being advanced is problematic/unproblematic, or as supposing that the reader needs to be convinced of a particular viewpoint (Martin and White 2005: 95). In the case of jointly-constructed texts, the editor assumes that the reader shares with him/her certain amount of knowledge to understand the requirements of the texts. In Example (7), for instance, the reader should have at least a general idea of the contour of human face, as well as the English vocabulary concerning the basic facial features. As for Examples (8) and (9), the English language level of the intended reader is assumed to be advanced enough to comprehend and perform the tasks.

However, the issue of alignment or solidarity is by no means a matter of total agreement between the communicative parties. As Martin and White point out,

…solidarity can turn, not on questions of agreement/disagreement, but on tolerance for alternative viewpoints, and the communality into which the writer/speaker aligns the reader can be one in which diversity of viewpoint is recognized as natural and legitimate.  (Martin and White 2005: 96)
The blanks or gaps in the jointly-constructed texts mentioned above are left for the reader to complete, while the reader still has certain degree of freedom, which depends largely on the context, to provide alternative answers other than the keys stipulated by the editor. Take the three examples in our data for instance. The reader of Example (7) is allowed to draw the human face in ways other than the finished one that is held by the character Mike. For example, the reader can depict the eyes with apparent pupils, or the mouth with teeth visible, or even a crying human face instead of a smiling one, as long as the depiction complies with the context of face-drawing. In the Ic PAIRWORK exercise of Example (8), the reader is free to use other wording to make the conversations, and yet this verbal communication should be concerned with the characters’ activities on vacation and should keep to the genre of dialogue. The reader of Example (9) enjoys more freedom in composing the speech draft, as long as the writing agrees with the communicative context of accepting the invitation from Mr. Zhangsha to give a talk about the stamps to honor the city.

Based on the orientation of reading position, Martin and Rose (2007a: 312-313; see also Martin and White 2005:62) classify different types of reading into three categories, i.e. compliant reading, resistant reading and tactical reading. Compliant readings take up, or subscribe to the reading position which is naturalized by the general trajectory of meanings in a text; resistant readings positions work against the naturalized reading position; while tactical readings rework the reading position naturalized in a text for some specific interests or social purposes. It can be inferred from the analyses above, that the reading position construed in the jointly-constructed texts in multimodal EFL textbooks is generally a compliant one, in that the completion of the texts requires the cooperative performance of the ideal reader. This dialogic editor-reader alignment permits to some extent the possibility of answers other than those prescribed by the editor. Nevertheless, the construed putative reading position only allows for the possible alternatives that comply with the given communicative context.
4.3.4 Illustration

According to Barthes (1977: 38-41), text-image relations can be classified into three types, i.e. ‘anchorage’ (text ‘elucidates’ image), ‘illustration’ (image elucidates text), and ‘relay’ (text and image ‘stand in a complementary relationship’). The major social purpose and function of EFL textbooks is for language teaching and learning, and there are quite a number of images clarifying or supporting the corresponding texts. The present section will be devoted to the discussion of the ENGAGEMENT meanings that are encoded in the images of illustration.

4.3.4.1 Image as the link between verbal texts

I shall take Example (10) for instance to explain how an image illustrates verbal texts by serving as the link between them. Example (10) is an excerpt from Unit 3 How many in the PEP Primary English Students' Book II for Year 3 (2003: 26). What is supposed to be learnt in this teaching section is the letters j, k as well as the words jeep, jump, kangaroo, and key. Here two images in comic style play an essential role in associating these otherwise unrelated words with each other in a meaningful way. To be specific, the image on the left links jeep with jump by describing a jeep frightening a fog so it jumped away; whereas the one on the right associates kangaroo with key by depicting a big key in a kangaroo’s pouch. In doing so, the words to be learnt are meaningfully connected through the funny cartoons, which is supposed to facilitate understanding and memorization. The editor tries to associate the words by resorting to cartoon characters, and the characters on the other hand acknowledge and lend support to editor voice. The degree of openness in the multimodal text is therefore increased. In this sense, the illustration functions to realize the ENGAGEMENT meaning of [attribute].
4.3.4.2 Illustration of improper behaviours

The majority of the characters in the EFL textbooks are examples of proper behaviours. Nonetheless, illustrations of improper behaviours can still be found in these textbooks. For example, Example (11) below is taken from Unit 1 Our school in the *PEP Primary English Students’ Book II for Year 4* (2003: 12).

In Example (11) editor voice is explicit in the five imperative clauses, which advise the reader not to violate the regulations in school. Four of the five imperative clauses proscribe the violations by using negative polarity, i.e. *Don’t drink or eat in the computer room*, *Don’t walk on the grass in the garden*, *Don’t push in the hallway*, and *Don’t waste food in the canteen*. As for the visual displays, five images are employed to illustrate the imperative clauses in the verbal texts by describing five instances of violation, i.e. playing noisily in the library, drinking and eating in the computer room, walking on the grass in the garden, pushing in the hallway, and wasting food in the canteen. It can be inferred that in Example (11) editor voice in the verbal texts and character voice in the five illustrative images go against each other. In between these five illustrations, there is another small image depicting an anthropomorphized fox holding the billboard and shouting *NO*, which further strengthens the meaning of denial. The illustrative images here functions to contract
or close down the heteroglossic space, and consequently the degree of openness to other voices is reduced. Therefore, it is argued that the five illustrative images in Example (11) encode the ENGAGEMENT meaning of [disclaim].

(11)

![Example (11) Illustrations]

[PEP Primary English Students’ Book II for Year 4 2003: 12]

4.3.4.3 **Background and foreground illustration**

There is still another way of lending support to editor voice through the use of illustration, in which the supportive image is backgrounded or foregrounded. A case in point is Example (12), which is taken from *Unit 1 Friendship* in the senior secondary EFL textbooks *New Senior English for China Student’s Book 1* (2004: 2).

In Example (12) the verbal text entitled *ANNE’S BEST FRIEND* tells the story about a Jewish girl named *Anne Frank*, who had to hide away from German Nazis in World War II. She treated her diary as her best friend that she called *Kitty*, and recorded her experiences and feelings in it. At the end of the verbal text an imperative clause is used (i.e. *Now read how she felt...*) to direct the reader to read an extract of *Anne’s* diary, which takes the form of a piece of torn paper from the diary. A monochromatic photograph of *Anne* is presented as the background of the first
paragraph of the verbal text *ANNE’S BEST FRIEND*, where the character *Anne* is first introduced. *Anne*’s voice is further detected in the image of the extract of diary, at the foot of which the sepia-tone photograph of her diary is foregrounded. The photographs of *Anne* and her diary represent *Anne* as a real person and her diary as a tangible genuine document, which bear evidence of the facticity of the anecdote. Here the voice of the character *Anne* lends support to editor voice in the verbal text via illustration, which functions to expand the heteroglossic space and can be identified as the ENGAGEMENT device that encodes [attribute]. Meanwhile, through supporting editor voice, this attribution further enhances the authenticity of what is described by the editor in the story, hence reinforcing the ENGAGEMENT meaning of [proclaim] in the text.

(12)

![Anne's Best Friend](image)

*New Senior English for China Student’s Book 1 2004: 2]*

### 4.3.5 Highlighting

Certain elements in a visual display are more eye-catching than the others, hence
creating a hierarchy of visual importance. According to Kress and van Leeuwen (2006: 201-203; see also Jewitt and Oyama 2001), the high degree of visual importance is referred to as ‘salience’, which can be realized through relative size, color contrast, tonal contrast, placement in the composition, sharpness of focus, and all other means that attract the viewer’s attention. In this section I shall examine a way of prioritizing certain visual elements via the choice of colour or typeface, i.e. highlighting in visual image and verbal text, and examine the way it contributes to the construction of a heteroglossic setting.

4.3.5.1 Highlighting in visual image

I shall begin the investigation with the analysis of the highlighting in visual images. Take Example (13) for instance, which is taken from Unit 2 Where is the Science Museum in PEP Primary English Students’ Book I for Year 6 (2003: 16). The whole image includes two parts. The upper part depicts a girl asking a policeman about the location of the library. The word library is highlighted in the dialogue balloon that conveys the voice of the girl.

(13)

![Image of the dialogue balloon with "library" highlighted in blue]

In the dialogue balloon, the word library is highlighted in blue, which indicates that it
is merely one of the many possible options that can be chosen from. Other alternatives can be found in the five smaller images in the lower part, which depict five buildings, i.e. post office, hospital, cinema, bookstore and science museum. Each of these five alternatives can be substituted for the word library which has already been chosen in the dialogue balloon. If put into words, the ENGAGEMENT meaning of the reason for highlighting the word library is that ‘It is possible that you want to go to the library’ or ‘You may want to go to the library’. In addition, it is noteworthy that these alternatives all come from the multimodal discourse itself. In other words, the multimodal text per se has already determined the range of possible alternatives, i.e. the five locations presented in the five small images. Based on the analysis above, it can be argued that the highlighted word in Example (13) indicates the status of the highlighted word as one of the possible options against the wider backdrop that consists of alternatives from the multimodal text itself, hence functioning to realize the ENGAGEMENT meaning of [entertain]. The highlighted word library opens up the heteroglossic space to allow for other voices that come from the text itself instead of from somewhere outside the multimodal text.

### 4.3.5.2 Highlighting in verbal text

As was mentioned in Section 4.3.1 above, it is mainly the verbal text, rather than visual image as is in the primary and junior secondary EFL textbooks, that provides the context of language teaching in the senior secondary EFL textbooks. Accordingly, there must be other types of multimodal ENGAGEMENT device besides the practice of labelling that are adopted in the senior secondary EFL textbooks to negotiate meanings with and around character voice. In the current section I shall take the extract of Anne’s diary in Example (12) for instance to describe how the highlighting in verbal text function to mediate heteroglossic space.

The extract of Anne’s diary, as is represented in Example (14) below (bold-face in the original excerpt from the textbook), is written in the first person, as if Anne was sharing her experiences and feelings with her friend Kitty in a letter.
Thursday 15, June, 1944

Dear Kitty,

I wonder if it’s because I haven’t been able to be outdoors for so long that I’ve grown so crazy about everything to do with nature. I can well remember that there was a time when a deep blue sky, the song of the birds, moonlight and flowers could never have kept me spellbound. That’s changed since I was here.

… For example, when it was so warm, I stayed awake on purpose until half past eleven one evening in order to have a good look at the moon for once by myself. But as the moon gave far too much light, I didn’t dare open a window. Another time some months ago, I happened to be upstairs one evening when the window was open. I didn’t go downstairs until the window had to be shut. The dark, rainy evening, the wind, the thundering clouds held me entirely in their power; it was the first time in a year and a half that I’d seen the night face to face …

Yours,
Anne

The words and expressions highlighted in bold-face type signals the intrusion of textbook editor on this personal record of daily experiences/letter to a close friend. In other words, what are highlighted in the verbal text are the vocabulary and expressions selected by the editor for the students to learn. What are prescribed as the teaching content are emphasized by means of highlighting. It may be argued that the resource of highlighting here functions to expand the heteroglossic space and bring in editor voice, and hence can be identified in the ENGAGEMENT system as encoding [attribute].

4.3.6 Gradability of ENGAGEMENT values in multimodal discourse

As was reviewed in Chapter 3, gradability is a general feature of the ENGAGEMENT system. Based on the analysis above, I am now in a better position to discuss how the aforementioned multimodal meaning-making resources up-scale and down-scale ENGAGEMENT values. As was pointed out by Martin and White (2005: 135), the meanings that are scaled within the ENGAGEMENT system vary from sub-system to sub-system. Therefore, I concentrate the discussion here on the same type of ENGAGEMENT device that encodes the same ENGAGEMENT meaning.

As was analyzed in Section 4.3.2 above, dialogue balloons can be categorized into three types according to their functions: (i) those that support editor voice (see
Example (3)), (ii) those that explain rules of games by demonstration (see Examples (4) and (5)), and (iii) those that give directions to the reader (Example (6)). In the case of Type (i), editor voice is present in the communicative context, which is represented in a separate paragraph above the image of the two characters whose conversation lends support to editor voice. The support from character voice further enhances editor voice by providing evidence for the editor’s statement. The editor’s ‘proclamation’ is accordingly strengthened, and hence the heteroglossic space is in a sense contracted as compared with the two other types of dialogue balloons. Editor voice in Type (ii), on the other hand, is absent from the demonstration of games (but sometimes appears in the instruction), while character voice is responsible for the verbal and visual demonstration. In Type (iii), the editor who should have given directions chooses to hide away from the dialogic setting, and character voice is solely responsible for the exercise instruction. Therefore, it may be inferred that Type (iii) opens up more space for character voice than Type (ii), because the responsibility of giving instructions should have been undertaken by an editor whereas demonstrations are not necessarily given by an editor. Type (ii) in turn expands more heteroglossic space than Type (i), in that editor voice is absent from Type (ii) but present in Type (i), and character voice in Type (i) actually enhances editor voice. In other words, within the ENGAGEMENT device of dialogue balloon, GRADUATION operates along the axis of FORCE, that is, the amount of responsibility that the characters undertake in instructing the reader or the amount of space opened up to engage character voice. In terms of the locations of these three types of dialogue balloon along the gradability cline of [attribute] value, it may be summarized that Type (iii) is scaled as occupying the point with the highest value among the three, Type (ii) is situated at the point with medium value, and Type (i) is the one with the lowest value among the three, as is diagrammatically presented in Figure 4-2.
Based on the analysis in Section 4.3.3, jointly-constructed texts in the EFL textbooks can be subdivided into three types as well: (i) jointly-constructed drawing exercise (see Example (7)), (ii) multimodal jointly-constructed herald page (see Example (8)), and (iii) jointly-constructed verbal text with accompanying images (see Example (9)). The gradable [attribute] values encoded in jointly-constructed text can be approached in light of the degree of its completion or ‘fulfillment’ (Hood and Martin 2007), that is, its prototypicality (i.e. FOCUS) as a co-construction that involves reader’s participation. It may be argued that the higher degree of completeness a jointly-constructed text possesses, the less heteroglossic space it opens up to alternative voices. This is because a jointly-constructed text will not have much room for the external voices to participate in the co-construction if it is almost complete on its own. The image in Type (i) and the verbal text in Type (iii) are far from being finished. Moreover, the reader’s participation in drawing facial features or filling in the gaps adds on new meanings to the original text. These two types of jointly-constructed text invites, or even demands participation of the external reader voice, and therefore they may be scaled as possessing relatively high [attribute] value. Type (ii), on the contrary, is itself a complete image as far as drawing goes, and the square blanks that demand reader’s participation are later added into the complete image. In other words, the incompleteness is not an inherent property but a later-added feature in Type (ii), and hence it encodes low [attribute] value (See Figure 4-3 below).
Among the three types of illustration that are identified in Section 4.3.4, two of them encode the same type of ENGAGEMENT meaning of [attribute]: (i) image as the link between verbal texts (see Example (10)), and (ii) background and foreground illustration (see Example (12)). The FORCE of [attribute] meaning that is realized in illustration can be investigated by looking at the indispensability of the image in linking different parts of a verbal text. As was analyzed in Section 4.3.4 above, Type (i) functions to associate the two semantically unrelated words with each other through the use of a meaningful and intriguing visual display. In other words, there would not be any semantic relevance between the two words with the same initial letter if the image were removed. As compared with Type (i), the de-focused, backgrounded photograph of Anne and the foregrounded photograph of Anne’s diary in Type (ii) is not so indispensable to the whole multimodal text. Even if the image were removed, the verbal text would still be coherent and consistent. Therefore, it may be concluded that Type (i) calls for the alternative voice of the characters to integrate different parts of the verbal text, hence opening up more heteroglossic space. Accordingly, Type (i) can be graded as encoding higher [attribute] value than Type (ii), which does not play an essential role in linking the verbal text, as is demonstrated in Figure 4-4.

Figure 4-4 FORCE of two types of illustration in realizing [attribute] value
As was stated at the beginning of this section, the current discussion of the gradability of ENGAGEMENT values is mainly conducted within the same multimodal ENGAGEMENT devices that encode the same meaning in the EFL textbook discourse. Further studies may be conducted to find out other multimodal ENGAGEMENT devices in various communicative contexts, and a GRADUATION continuum may be established within the same or across different semiotic resources that encodes the same ENGAGEMENT meaning.

4.4 Voice interaction in multimodal EFL textbook discourse

In the previous section, I have identified five types of multimodal ENGAGEMENT devices in EFL textbook discourse (i.e. labelling, dialogue balloon, jointly-constructed text, illustration and highlighting), and accounted for how these multimodal resources realize the ENGAGEMENT meanings of [disclaim], [proclaim], [entertain] and [attribute]. These multimodal ENGAGEMENT devices contribute to the expansion and contraction of the heteroglossic space in a text. The present section will be concerned with the interplay of the multiple voices, and will also explore the interactive meaning encoded in multimodal EFL textbook discourse in terms of contact/observe, social distance and point of view (for the theoretical account see Section 3.4.2).

4.4.1 Focalization and eye contact

A large number of the images in the EFL textbooks depict people or human-like creatures with eyes and facial expressions, while the majority of these animate participants in the picture frame do not gaze at the viewer. In other words, they are ‘observe’ images (Painter 2007; see Section 3.4.2 for further discussion). Some of these represented participants in our data are characters in a story (e.g. Examples (1),(2) and (3)); other provide demonstrations for the reader to follow (e.g. Examples (4) and (5)). The reader views them impersonally as items of information and objects of observation. This observation mainly happens directly instead of
vicariously. One notable exception is Example (1), in which the observation is mediated through the eyes of the focalizing characters *Amy* and her mother. In this image, the viewer is visually placed behind *Amy’s* mother, observing her as well as what she sees in *Amy’s* bedroom. By the same token, the viewer is again observing the wardrobe ‘along with’ the character *Amy*.

Figure 4-1 and Example (6) are exceptions to the dominant ‘observe’ type of image. In these cases, the represented participants look directly at the viewer, and therefore the viewer is engaged in an eye contact with the characters. These represented participants are requiring the reader to do the exercises or answer questions according to the requirements, and their gestures and/or the utterances in dialogue balloons further specify what is required. In other words, the represented participants are portrayed as people directly addressing the reader instead of passive phenomena merely for observation. These images belong to the ‘contact’ type, and more often than not they are also jointly-constructed texts that demand reader’s participation. Together with the imperatives and interrogatives in the dialogue balloons, the ‘contact’ images invite the reader to participate in the jointly-constructed texts, and hence bring in reader voice and encode the ENGAGEMENT meaning of [attribute].

### 4.4.2 Reader involvement and social distance

The choice of social distance is another parameter that indicates various types of interaction between the viewer and the represented participants inside the picture frame. In visual communication, the different degrees of social distance are interpreted as the continuum of the ‘size of frame’ of shots (Kress and van Leeuwen 2006: 124). In the present data of EFL textbooks, the images cover a wide range of shots from ‘extreme close-up’ to ‘long shots’, which suggests the extent of distance variations from ‘intimate distance’ to ‘public distance’ with many intermediate degrees.

The social distance implied in the visual displays corresponds with the voice interaction. For instance, the image of Example (13) shows the full figures of the
represented participants and their surroundings, and thus the relationship between the viewer and the represented participants is that of ‘public distance’. As is revealed in the analysis in Section 4.3.5.1 above, the ENGAGEMENT meaning of [entertain] is encoded in this image through the use of highlighting, which indicates that alternative positions all come from the internal source within the multimodal discourse itself. In other words, the reader is not involved in determining the range of possible options, because they are already included in the five small images in the lower part of the multimodal text. Therefore, it may be justified to say that the distant, impersonal public distance complies with the ENGAGEMENT meaning of [entertain].

Public distance is also adopted in those images in which editor voice engages with character voice (such as Examples (1), (2) and (11)). In these cases, the reader is addressed by the proclamation or disclamation from editor voice instead of actively engaging himself/herself with the voices within the discourse. This low degree of reader’s involvement also echoes the choice of public distance. The high degree of reader’s involvement can be found in jointly-constructed texts (e.g. Figure 4-1, as well as Examples (6) and (7)). The medium shots that show the figure from waist up in Figure 4-1 and Example (6) suggest the ‘far personal distance’, while the extreme close-up in Example (7) implies ‘intimate distance’. The represented participants in these images are presented within easy touching distance to the viewer. This agrees with the meaning of [attribute] encoded in these jointly-constructed texts, in which the reader is invited to participate in the ultimate completion of the texts and thus has access to the field of heteroglossia.

Nevertheless, the discussion of the association of voice interaction and social distance here is by no means claiming that voice interaction is a decisive role in choosing the patterns of social distance. The choice of shots that conveys the degree of social distance is a complex issue. Other factors such as the conventions in visual genres (see Kress and van Leeuwen 2006: 126) should also be taken into account before a comprehensive understanding of the social distance in visual images is achieved.
4.4.3 Character engagement and point of view

The third factor that should be considered in relation to voice interaction is point of view, which can be approached along the horizontal axis and vertical axis. As was reviewed in Section 3.4.2 in Chapter 3, the horizontal angle conveys whether the interactive participant (i.e. image producer and viewer) is involved with (i.e. frontal angle) the represented participants or is detached from (i.e. oblique angle) from them; while the vertical angle implies whether it is the interactive participant has power over the represented participants (i.e. high angle) or vice versa (i.e. low angle) or their relation is a equal one (i.e. eye level angle) (Kress and van Leeuwen 2006: 133-143).

The majority of the images under examination adopt the frontal and eye level angle, and hence what is depicted in the image is presented as ‘part of our world’. In other words, the relation between the represented participants and the viewer is one of equality. The frontal angle indicates a high degree of reader identification and involvement, which has a lot to do with the engagement of character voice. Although the power differences do exist between the textbook editor and reader due to the unequal possession of knowledge, the interpolation of characters into the heteroglossic setting enables the reader to negotiate alignment with the characters on an equal footing. That at least partially explains why the represented participants are seen from the point of view of the eye level, rather than from a low angle to demonstrate power to the viewer.

In closing, I attempt to summarize the interaction between editor voice, reader voice and character voice in the multimodal EFL textbook discourse in Figure 4-5 below. The functions of the multimodal ENGAGEMENT devices (i.e. labelling, dialogue balloon, jointly-constructed text, illustration, and highlighting) are indicated above or underneath the arrows between the multiple voices to show how these voices interact with one another. To be specific, the broken-line arrow between editor voice and reader voice indicates that editor voice overtly instructs the reader through the use of section titles that may take the forms of imperatives or nominal groups, whereas the real-line arrow represents the way in which reader voice interpolates editor voice via jointly-constructed text. The arrows between editor voice and character voice
show that dialogue balloon and illustration introduce character voice into the otherwise monoglossic texts, while editor voice intrudes itself into character’s visual story or verbal recount by attaching labels or highlighting words in verbal texts. As for the interaction between character voice and reader voice, a character may undertake the responsibility of instructing the reader via dialogue balloon, whereas reader voice is engaged in the co-construction of multimodal text in which character voice more often than not provides the visual communicative setting. It may be inferred that the five types of multimodal ENGAGEMENT devices, along with other semiotic resources, work in concert to construe the heteroglossic harmony in EFL textbook discourse.

Figure 4-5 Voice interaction in multimodal EFL textbook discourse (Chen and Qin 2007)

4.5 Summary

In the present chapter, I have first analyzed and discussed the multimodal resources for the management of heteroglossic space in multimodal EFL textbook discourse. Five types of multimodal resources have been identified as the ENGAGEMENT devices that encode the ENGAGEMENT meanings of [disclaim], [proclaim], [entertain] and [attribute]. To be specific, the practice of labelling in multimodal texts signals the intrusion of editor voice, thus encoding the ENGAGEMENT meaning of [proclaim]. The use of dialogue balloon and illustration brings in character voice, and a given
proposition or viewpoint is thus attributed to the character(s). Jointly-constructed text opens up the space for dialogic alternatives involving the reader, and contributes to the construal of an ideal or putative readership. Highlighting functions to entertain other alternatives or possibilities grounded within the contingent subjectivity. Although all the images that contain multimodal ENGAGEMENT resources have not been displayed in the current analysis, the aforementioned five types of ENGAGEMENT devices are commonly found in the EFL textbooks under discussion.

Acknowledging the fact that gradability is generally a property of the ENGAGEMENT system, I have investigated how the multimodal devices alluded to above scale up or down the ENGAGEMENT values encoded in the multimodal EFL textbook discourse. It is shown that the three types of dialogue balloon can be located along the gradability continuum of [attribute] values based on how much responsibility the characters undertake in construing the heteroglossic setting. In light of the degree of completion, the three types of jointly-constructed text open up heteroglossic space to different extents for the external reader voice to participate in the construction of the text. As for the ENGAGEMENT device of illustration, the FORCE of the [attribute] meaning can be investigated from the perspective of the indispensability of the image in linking the different parts of the corresponding verbal text.

After the exploration of the heteroglossic nature of the multimodal EFL textbook discourse, a discussion concerning the voice interaction in relation to contact/observe, social distance and point of view has also been provided in this chapter. It is found that the characters’ eye contact with the reader is only established in those images where the characters give directions and require actions from the reader. The different choices of social distance indicate various types of relationship between the characters inside the picture frame and the reader who observes them. Moreover, the interpolation of characters accounts for the adoption of the frontal and eye level angle in most of the images. Based on the investigations above, a triangle involving editor voice, reader voice and character voice is set up, which demonstrates the roles of the multimodal ENGAGEMENT devices in the play of multiple voices in multimodal EFL
textbook discourse.

So far the focus of the discussion has been on the interactive aspect of the interpersonal management. However, as was pointed out in Chapter 3, interpersonal semantics also include the personal dimension. The ensuing two chapters will address this issue from two different perspectives respectively, i.e. the evaluative meanings encoded and visual modality.

NOTES

1 Considering the fact that the main purpose and function of EFL textbooks is for language teaching, I use ‘reader’ in the term ‘reader voice’ to refer to both reader of verbal texts and viewer of visual images. Nevertheless, the term ‘viewer’ is sometimes adopted in the current research when the analysis mainly centres around visual images (e.g. Sections 4.3.2.3 in this chapter).

2 The labelling in Figure 4-1 is different from labelling in the general sense, in that it contains a missing letter that is intended to involve reader’s participation. It could be termed as ‘jointly-constructed labelling’, which demonstrates features of both labelling and jointly-constructed text. Accordingly, the ‘reader voice’ and ‘editor voice’ indicated in the labelling in Figure 4-1 as a matter of fact represent a dual voice that implies editor’s closer engagement with reader at the early stage of English learning.

Chapter 5

Attitudinal Accumulation:
Verbiage-image Complementarity and Co-instantiation

5.1 Introduction

In the previous chapter, I concentrated the discussion on the interactive dimension of interpersonal management in EFL textbooks. This chapter will be devoted to the personal aspect, to be precise, the evaluative meanings encoded in multimodal discourse. The intermodal relation between image and verbiage in encoding evaluative meanings has warranted scholarly attention. As is pointed out by Martin (2001: 334), ‘as far as interpersonal meaning is concerned, verbiage/image relations are more concerned with APPRAISAL than mood or modality’. The present study aims at examining the ways in which verbal and visual resources are deployed to encode attitudinal meanings, with particular attention paid to the verbiage-image relations. The analytical frameworks drawn upon here are the semantic system of APPRAISAL (Martin 2000; Martin and White 2005), and systemic functional semiotic approach to multimodal texts (Kress and van Leeuwen 2006). Based on the studies the ontogenesis of ATTITUDE system (Painter 2003), the current study will conclude by attempting to account for how attitudinal meanings are institutionalized and accumulated as students progress from primary to secondary schooling.

5.2 Attitudinal meanings in multimodal discourse

As was reviewed in Chapter 3, the system of ATTITUDE covers three semantic
dimensions, i.e. AFFECT, JUDGEMENT and APPRECIATION. In traditional, non-technical terms, they are respectively referred to the inborn human emotions or feelings, the ethical evaluation of people’s behaviours, and the aesthetic awareness of man-made things or natural phenomena. The present study endeavours to examine the verbiage-image relations in encoding the attitudinal meanings within multimodal EFL textbook discourse. Therefore, the social semiotic approach to verbiage-image relations will be briefly reviewed, and the analytical tools for describing verbal and visual attitudinal meanings will be outlined.

5.2.1 Social semiotic approach to verbiage-image relations

The verbiage-image relations have arrested scholarly attention from social semioticians over the past decade. As is pointed out by Martinec (2005: 166), the studies on verbiage-image relations include both the relations between different componential parts of a multimodal text, and the relations between two or more whole multimodal texts. The first category can be further divided into the studies on the textual, ideational and interpersonal dimensions.

5.2.1.1 Relations between components of a multimodal text

In their foundational work on multimodal discourse, Kress and van Leeuwen (2006) focus on the textual aspect of verbiage-image relations. As was reviewed in Chapter 3, the co-authors study the Given-New, Ideal-Real and Centre-Margin composition of verbal texts and visual images. Martin (2001, 2004) suggests expanding this framework to include Theme/Rheme, in that an image Theme can function as an interpersonal Theme (Halliday 1994), and hence can be deployed to naturalize the evaluation of the ensuing verbal Rheme.

Concerning the ideational dimension of verbiage-image relations, Lemke (1998a) investigates the interaction of different semiotic systems, i.e. language, tables, graphs, images and diagrams, in modern science texts. He points out that the interplay of these multiple semiotic systems creates new orders of meaning (‘multiplying meaning’ as he puts it), and thus recontextualizes the scientific knowledge.
O’Halloran (1999a) studies the meaning arising from the interaction and interdependence between language, visual display and mathematical symbolism, with a particular focus on ‘semiotic metaphor’, which is referred to the phenomenon of metaphorical representation across semiotic codes. The intersemiotic semantic relations between visual and verbal components of a text are approached by Royce (1998, 2002) through the concept of ‘intersemiotic complementarity’. Drawing upon the study on lexical cohesion (Halliday and Hasan 1976), Royce identifies six ideational intersemiotic sense relations, i.e. repetition, synonymy, antonymy, hyponymy, meronymy, and collocation. Still mainly concentrating on the ideational dimension, Martinec and Salway (2005) combine Halliday’s (1994) logico-semantic relations (elaboration, extension, enhancement and projection) with Barthes’ (1977) classification of image-text relations to build a generalized system of image-text relations. They also identify the units and realizations of both the logico-semantic and status relations, and try to specify some machine-recognizable realizations. In most recent advances, O’Halloran (2008) proposes the development of new theoretical and practical approaches to multimodal discourse analysis by using digital technology in the form of image-editing software. Through the intersemiotic analysis as well as semantic and ideological interpretation, O’Halloran reveals how the metaphorical constructions of ideational meaning happen across linguistic and visual components in a print advertisement.

As for the interpersonal aspect of verbiage-image relations, Martin (2001, 2004) points out that APPRAISAL can be considered as a resource for relating image to verbiage, in that visual images in multimodal texts provoke the evaluation in the reader as imagery does, hence functioning to co-articulate ATTITUDE with the verbiage. Inspired by Martin’s (2001, 2004) studies, the present study takes the evaluative orientation of verbiage-image relations as the point of departure, and examines the co-instantiation and complementarity between the semiotic modes of visual images and verbal texts in encoding the attitudinal meanings within the multimodal EFL textbook discourse.
5.2.1.2 Relations between whole multimodal texts

In addition to the studies on the relations between different componential parts within a multimodal text as was briefly reviewed above, the social semiotic approach to verbiage-image relations also includes the relations between whole multimodal texts, each of which is created in one semiotic mode. Iedema’s (2001, 2003) concept of ‘resemiotization’ belongs to this category. He shows us that different semiotic modes are linked on the basis of a resemiotisizing logic, and it is through the process of resemiotization that the knowledge and technologies of a community, as well as its interpersonal and social practices are transposed and reified. Other studies that explore the possibility of the ‘translation’ between different semiotic modes include Martinec and van Leeuwen’s (2008) work on translating different text types into different kinds of diagrams. By categorizing the various text types and diagram types based on such relations as centre-margin and classification, they identify two units of relations and entities, and establish the equivalence between the realizations of these two units in texts and diagrams.

The present study falls into the category of the research on the first type of verbiage-image relations, i.e. the relations between componential parts of a multimodal text. Before examining the verbiage-image relations from the attitudinal perspective, I shall first outline the framework for analyzing ATTITUDE in verbal text and visual images.

5.2.2 Inscribed and invoked ATTITUDE

In SFL, attitudinal meanings have first been explored in verbal texts before the discussion extends to attendant modes such as visual displays. The realizations of ATTITUDE cover the overtly construed attitudinal meanings, and the implicit implications designed to arouse evaluation. These two types of resources are technically termed as inscribed ATTITUDE and invoked ATTITUDE respectively.

5.2.2.1 Verbal inscription and invocation

In the tri-stratal system of language, ATTITUDE is realized through explicit or implicit
means at the level of lexicogrammar (Macken-Horarik 2003; Martin 2000a; Martin and White 2005; Page 2003). It can be directly inscribed in text via the use of evaluative lexis. For example, the attitudinal meanings can be encoded in epithet as in a magnificent palace, or in the attribute of a relational process such as the servant is honest, or in the circumstantial elements like the child went home happily.

In addition to the overt realizations, attitudinal meanings can also be indirectly invoked by using a range of strategies. Martin and White (2005: 61-68) explore the ways of invoking attitudinal meanings, including the use of lexical metaphor in provoking an attitudinal response from the reader, the use of non-core vocabulary infused with manner and counter-expectancy to flag or connote a positive or negative orientation to the proposition that is advanced in discourse, and the deliberate selection of ideational meanings in affording an ATTITUDE. Martin and White (2005) provide an outline of the strategies for inscribing and invoking ATTITUDE in verbal texts, as is presented with examples in Figure 5-1. If the cline is read in a top-down manner, the increase in the degrees of freedom that allows the reader to align with the naturalized value positions in the text can be perceived.

Figure 5-1 Verbal strategies for inscribing and evoking ATTITUDE (Martin and White 2005: 67)

5.2.2.2 Visual inscription and invocation

Recent studies on the APPRAISAL system have been extended to cover the visual mode. The pioneering research in this area reveals the attitudinal repertoire in visual images,
as well as the role of images in construing the global evaluative stance in a multimodal text.

The issue of invoked ATTITUDE in visual images has been approached by analysts who study media texts. Economou (2006) examines the verbiage-image unit of ‘lead image-headline’ in weekend news review feature stories from a Greek newspaper and an Australian newspaper. Drawing on Kress and van Leeuwen’s (2006) framework for visual analysis and the APPRAISAL system (Martin 2000a; Martin and White 2005), she argues that all three types of attitudinal meanings (i.e. AFFECT, JUDGEMENT, and APPRECIATION) can be ‘provoked by choices in content, composition, and interaction’ in images. The choice of visual arrangements, as Economou (2006) explains, complies with the ‘desired readership’ and the ‘known political position’ of the particular newspaper.

As for the attitudinal inscription in visual images, it is evident that AFFECT can be inscribed through the depictions of emotive facial expressions. Tian (2007) describes how the affectual meaning of worry is inscribed in the children’s picture book *Silly Billy*. Her study demonstrates that the frequent use of ‘worry’ in verbal texts and its various realizations in images combine to produce evaluative meanings in multimodal texts. White (2006) contends that AFFECT is by no means the only attitudinal meaning that can be inscribed visually. Based on his observations on media images, White points out that both the presence of the image creator and the positioning of the viewer should be taken into consideration. If the image creator’s intentional manipulation is obviously signaled in the construction of an image and this manipulation is intended to put the viewer in the position to provide either a JUDGEMENT or an APPRECIATION, it is justified to categorize the image as inscribing JUDGEMENT or APPRECIATION. According to White (2006), in media discourse the manipulation may include the distortions in cartoons and caricatures, as well as the deliberate selections of certain photographs or portraits in photojournalism. As White explains, in these visual instances the image creator’s purpose is overtly demonstrated, and hence a certain JUDGEMENT or APPRECIATION is inscribed.

So far few attempts have been made to describe how the visual displays in
pedagogic materials inscribe or invoke personal emotion, ethical judgement and aesthetic appreciation. Accordingly it is appropriate to further explore these three aspects of attitude in educational context where the cultivation of proper attitudes and values is part and parcel.

5.3 Co-instantiation and complementarity in encoding AFFECT and JUDGEMENT

I now begin the analysis of the attitudinal meanings encoded in multimodal texts focusing on the images and texts that describe people or anthropomorphized characters and their activities. The categories within the system of attitude that will be drawn on in the present section are AFFECT and JUDGEMENT. These two semantic dimensions provide powerful tools to describe the emotions emanating from participants endowed with consciousness as well as to evaluate the ways they behave or the character they possess. According to the different drawing styles, the images under discussion are classified into the categories of cartoon, portrait and photograph. These three styles are unevenly distributed among the EFL textbooks for different levels of education. As is indicated in Table 5-1, cartoon is the overwhelming style in the primary EFL textbooks and junior secondary EFL textbooks, accounting for 97.9% and 89.6% respectively. On the contrary, in the senior secondary EFL textbooks the numbers of portraits and photographs increase dramatically, with the proportion being 16.5% (as compared with 1.9% and 1.1% in the primary and junior secondary EFL textbooks) and 53.8% (in comparison with 0.2% and 9.3% in the primary and junior secondary ones). In this section portraits and photographs will be analyzed under the same heading. This is motivated by the fact that the represented participants in cartoons are portrayed with a certain degree of deliberate distortion, whereas portraits and photographs belong to the realistic style that aims at representing the participants as faithfully as possible. These human or personified participants either overtly convey the evaluative position, or indirectly invoke the attitudinal response in the reader.
Table 5-1 The distribution of visual styles in EFL textbooks for different levels of education

<table>
<thead>
<tr>
<th></th>
<th>cartoon</th>
<th>portrait</th>
<th>photograph</th>
</tr>
</thead>
<tbody>
<tr>
<td>primary textbooks</td>
<td>839 (97.9%)</td>
<td>16 (1.9%)</td>
<td>2 (0.2%)</td>
</tr>
<tr>
<td>junior secondary textbooks</td>
<td>403 (89.6%)</td>
<td>5 (1.1%)</td>
<td>42 (9.3%)</td>
</tr>
<tr>
<td>senior secondary textbooks</td>
<td>27 (29.7%)</td>
<td>15 (16.5%)</td>
<td>49 (53.8%)</td>
</tr>
</tbody>
</table>

5.3.1 Cartoon

According to The New Encyclopaedia Britannica, Volume 2 (2005: 910), a cartoon is defined as ‘originally, and still, a full-size sketch or drawing used as a pattern for a tapestry, painting, mosaic, or other graphic art form, but also, since the early 1840s, a pictorial parody utilizing caricature, satire, and usually humour.’ In the primary and junior secondary EFL textbooks under discussion, the cartoon is the most frequently adopted style. In spite of the fact that differences do exist between the cartoons in primary and secondary EFL textbooks, which will later be examined in Chapter 6 from the perspective of visual modality, the present chapter focuses on the attitudinal meanings that are encoded in this type of visual images with particular emphasis placed on the verbiage-image relations.

5.3.1.1 Inscribed happiness of English-learning

As is shown in Table 5-1, the cartoon is statistically dominant in the primary and junior secondary EFL textbooks. The majority of those cartoons in the EFL textbooks for primary education contain conscious participants that cover a wide range of cartoon characters from primary school students, their teachers and family members, to the personified animals endowed with sense and sensibility. Those cartoon characters that have a high frequency of appearance are listed in a separate page (see Appendix I) at the start of the textbooks. It is evident in Appendix I that most of the characters are primary school students with a similar age as the reader. Throughout the multimodal textbooks they are depicted as either learning English in class, or involved in everyday activities.
Example (1) is taken from Unit 1 How do you go there in the PEP Primary English Students’ Book I for Year 6 (2003: 5). It is typical of the cartoons that are widely used in primary EFL textbooks, which describe four school-age characters conducting a group discussion. Example (1) contains two cartoons in a successive relation to one another, comprising two shots of the scene in which four primary students conducting a group discussion. Each of the characters in the cartoons demonstrates an expressive countenance indicative of delight, which can be observed from the characters’ disproportionately large eyes with enlarged pupils that imply great interest, their facial expression of laughing from side to side with eyes closed, as well as their demonstrative and enthusiastic gestures. Positive attitudinal meaning, AFFECT to be exact, is thus directly inscribed in the visual display, with the school-age cartoon characters as Emoters (i.e. the conscious participants who experience the emotion) and the English-learning activities as the Trigger (i.e. the phenomenon that is responsible for that emotion) (Martin and White 2005: 46).

(1)

[PEP Primary English Students’ Book I for Year 6 2003: 5]

Drawing on Martin and White’s (2005) classification of AFFECT based on the six factors (see Section 3.3.1 for the detailed discussion), I analyze in detail the types of emotions that are encoded in Example (1). According to Martin and White (2005: 46-52), types of AFFECT can be categorized in light of the following six factors: (i) whether the emotions are construed as positive or negative in a given culture; (ii) whether they are realized as an emotional surge combined with behavioural
manifestations, or as an internal mental process; (iii) whether the feelings are reactions to a certain emotional Trigger, or as an undirected ongoing mood; (iv) how the emotions are graded along the clined scale ranging from higher value to lower value; (v) whether the emotions involve a reaction to a present or past stimulus, i.e. realis values, or involve an intention with respect to some prospective stimulus, i.e. irrealis; (vi) whether they are concerned with ‘affairs of the heart’ (un/happiness), ‘ecosocial well-being’ (in/security) or the ‘pursuit of goals’ (dis/satisfaction). The types of AFFECT in Figure 1 and their corresponding visual patterns can be analyzed and demonstrated diagrammatically in Table 5-2 below. It can be inferred from the analysis that the cartoons in Example (1) commit positive AFFECT through a series of overt, direct visual realizations (i.e. the depictions of facial expressions and behavioural manifestations).

<table>
<thead>
<tr>
<th>Types of AFFECT</th>
<th>Corresponding visual patterns</th>
</tr>
</thead>
<tbody>
<tr>
<td>positive AFFECT</td>
<td>cheerful facial expressions showing great interest in the ongoing activities in class</td>
</tr>
<tr>
<td>behavioural surge</td>
<td>extralinguistic manifestation of exciting gestures indicative of active engagement</td>
</tr>
<tr>
<td>reaction to other</td>
<td>feelings directed at the ongoing English-learning activity</td>
</tr>
<tr>
<td>high value</td>
<td>Overtly-expressed emotions, high-spirited fun and great merriment</td>
</tr>
<tr>
<td>realis AFFECT</td>
<td>emotions involve a reaction to the present stimulus of group discussion</td>
</tr>
<tr>
<td>happiness, security, satisfaction</td>
<td>appearance of comfort and enjoyment, confident when expressing themselves, absorbed and satisfied in the group work</td>
</tr>
</tbody>
</table>

As mentioned above, the two successive cartoons portray two moments when four primary students are practicing how to ask and answer questions concerning means of transport. The verbal texts in Example (1) are framed in four dialogue balloons, which demonstrate two dialogues between the students.

(Cartoon on the left)

_Wu Yifan:_ How do you go to school?
_Sarah:_ Usually I go to school on foot, because my home is near.
The verbiage is mainly concerned with the students’ experiences, i.e. how they usually go to school and the reasons for choosing certain transportation means. As for the interactive aspect of interpersonal management, the speech roles of demanding and giving information are enacted, and the speech functions of question and statement are realized by the relevant Mood choices. However, when it comes to the personal aspect (Thompson 1996: 68-69), attitudinal meanings are not found in the verbal texts, neither inscribed nor invoked. It is hard to tell simply by reading the verbal texts whether the students who utter the words are happy or upset with the language learning activity. It is equally difficult to nail down whether they are confident or frustrated when giving the explanations. In other words, the verbiage in Example (1) commits no attitudinal meanings. As a matter of fact, it is through the visual images, in the case of Example (1) the cartoons, that the affectual meanings are conveyed. The students feel excited, self-assured and engrossed, as represented in the images.

It may be helpful to take into consideration the function of visual demonstration of this type of images (see Section 4.3.2.2 for further discussion) so as to gain a better understanding of the verbiage-image relation in Example (1). As is shown in Example (1), the title of the teaching section is Group work, while the visual images vividly describe the scene of a group of students working together in class, with a mood of happiness added via visual affectual inscriptions. In a nutshell, the verbal texts within the dialogue balloons explain to the reader the way they are supposed to perform in the task by demonstrating of two model dialogues, whereas the images indicate the attitude that the reader is supposed to assume in undertaking that task, and perhaps in learning English in a broad sense. The atmosphere of happy English-learning is thus created in Example (1), and the verbiage-image relation in terms of attitudinal meanings is complementarity, in that the images commit positive affectual meanings that are inscribed in a range of visual resources which have no
counterparts in the verbiage. As was reviewed in Section 2.4.3 of Chapter 2 above, developing positive emotions and attitudes in students has been recognized as an essential part of the overall goal of the primary and secondary EFL education. As the current *Curriculum Standards for English* states (Zhonghuarenmingongheguo jiaoyubu 2001a: 1), it is a crucial task for EFL education to arouse and cultivate students’ interest in English learning as well as to build up their confidence in the learning process. The analysis of Example (1) above may account for how the multimodal meaning-making resources in EFL textbook discourse, especially the visual patterns, contribute to the unprecedented focus on attitude in EFL education in China.

5.3.1.2 *Gradability in affectual inscription*

*Attitude* involves gradable meanings, which means it has the potential to be intensified and compared (Martin and White 2005: 44). Feelings in other words have depth and attitudinal meanings can be graded. The gradability of attitudinal meanings in visual images can be investigated through the comparison between textbooks for different levels of education. Example (2) is extracted from *Unit 7 How do you make a banana milk shake* in the junior secondary EFL textbook *Go for it Students’ Book I for Year 8* (2005: 43), which is an instance similar to Example (1) in the sense that they both describe students’ group work activity.

Example (2) also involves two cartoons, portraying two teams of junior secondary students playing a *Recipe game*, with different background colours indicating the two different teams of students. Instruction on how to play the game is given in the verbal texts above the cartoons:

**GAME** Recipe game

Make two teams. Write a recipe. Then cut it up.
The other team has to put the recipe in order.
There are still two other verbal texts framed within oval bubbles, which demonstrate a sample recipe. The bubble on the left gives us a glimpse of the first part of the recipe that the first team of students write. The cartoon on the left correspondingly describes a team of three students busy writing the recipe.

First, cut up the tomatoes and beef. Next, boil the noodles.

The oval bubble on the right represents the remaining part of the recipe cut into pieces but in a correct order, while the corresponding cartoon beneath it depicts another team of four junior secondary students working together to piece together the recipe.

Then add the ingredients to the noodles.
Add salt to the noodles. Eat the noodles.

The verbal texts give directions and provide a model demonstration, while the cartoons work in tandem with the verbiage to visually represent the scene of game-playing. More importantly, the attitudinal meanings that are not conveyed in the verbiage are realized visually. Example (2) will be examined in comparison with Example (1) in terms of the scaling of FORCE in affectual inscriptions. Their similarity between the images of these two examples lies in the attitudinally committed visual images. The affectual dispositions inscribed in both images are
positive, directed at specific phenomenon. They convey happiness, confidence and contentment, and hence the positive ATTITUDE towards English learning is promoted. However, the represented participants in Example (2) are happy but less exuberant than those in Example (1). The feelings of comfort and enjoyment can be detected in their smiling and attentive faces, and yet the smile is quite faint in comparison with that in Example (1). Moreover, the gestures depicted in Example (2) are merely those that are necessary in playing the recipe game, whereas other possible gestures that may indicate the represented participants’ excitement and other joyful feelings are not found. In other words, a sense of seriousness is incorporated into the junior secondary textbook. When considering at what point the affectual inscriptions in Example (2) should be put along the cline of intensity ranging from high values to low values, it is justified to argue that the cartoons in Example (2) is scaled as less affectually committed than those in Example (1) in terms of the degree of the FORCE of the positive affectual meanings. If the AFFECT encoded in the cartoons of Example (1) is graded as high, it can be argued that the AFFECT inscribed in the cartoons of Example (2) is down-scaled towards the median value along the cline of affectual intensity.

As was dwelt on above, the AFFECT encoded in visual images is as gradable as that encoded in verbal texts. In other words, the options from the sub-systems of AFFECT and FORCE within the APPRAISAL system are ‘coupled’ (Martin 2008b; see Section 3.2.3 for a more detailed discussion) to different degrees. It may be fair to contend that the gradual changes in the degree of happiness demonstrated in the images of primary and junior secondary EFL textbooks reflect the process of children’s socialization, i.e. the transition from the playful and carefree childhood to adolescence, when a teenager has gone through puberty but not yet reached full maturity and thus a sense of responsibility should be cultivated. As will be shown in the analyses in Section 5.3.2, the earnest and non-jesting mood is more evident in the senior secondary EFL textbooks. The issue of emotion and attitude education in relation to the socialization process will be further discussed in Section 5.5 below, following the investigation into the invocations of JUDGEMENT and APPRECIATION in
visual images and the verbiage-image relations in encoding attitudinal meanings within the senior secondary EFL textbook discourse.

5.3.1.3 Logogenetic recontextualization

Now I turn to explore attitudinal dimensions other than AFFECT in the animate cartoons of multimodal EFL textbooks, with the focus still placed on the verbiage-image relations in terms of ATTITUDE -encoding. As was analyzed in detail in the previous section, the overwhelming majority of the images in primary EFL textbooks are inscribed with a high degree of positive affectual meanings. This section will examine the attitudinal response this type of images may invoke in the viewer, and show how the deployment of multimodal resources accords with the cultivation of moral standards that is advocated in the current curriculum standards in China. I shall take Example (3) for instance to exemplify the invoked ATTITUDE in the cartoons of EFL textbooks.

Example (3) is taken from Unit 2 My family in PEP Primary English Students’ Book II for Year 3 (2003: 15), and its main body is a cartoon depicting a family of five members. In the cartoon of Example (3), the feelings of the represented participants (i.e. the five cartoon characters) are made explicit via their emotive facial expressions, and hence the AFFECT of happiness is inscribed directly in the visual display. Nevertheless, in addition to the explicitly inscribed affectual meanings, the cartoon also invokes attitudinal response from the viewer through the way it presents the represented participants. The five represented participants, i.e. the five members in a family, are connected with each other by the vectors of arms and/or eyelines. To be specific, the right arm of the grandfather forms an oblique line, and thereby generates a vector linking him with his granddaughter. On the other hand, the eyelines emanating from the granddaughter and the father relate them to the grandmother and the mother. This visual connection, together with the inscribed happiness, presents to the viewer the scene of a united, harmonious and happy family get-together. In doing so, the image ‘invites’, or ‘affords’ to be more precise (Martin and White 2005: 67), a favourable evaluation of the filial piety and parental responsibility from the
putative viewer.

What is supposed to be invoked in the viewer of Example (3) is thus the positive APPRECIATION concerning the value of a family being united, through the JUDGEMENT of the ethical behaviour of filial piety and parental responsibility. In other words, The positive emotion is ‘reworked’ as ethics regarding propriety, which is in turn recontextualized as ‘politicized aesthetics’ (Martin 2002: 200) concerning the social value of harmonious family that conforms to the standard, conventional principle advocated in the given educational and social context. Adapted from Martin’s (2002: 199) model of logogenetic recontextualization process (cf. ontogenetic recontextualization to be discussed in Section 5.5.2) that is used to examine the role of evaluation in discourse of reconciliation, the attitudinal recontextualization in the cartoon of Example (3) can be diagrammatically represented in Figure 5-2 below. As Martin (2002) suggests, Lemke’s (1995) notion of ‘metaredundancy’ can be adopted to explain this recontextualization movement. As for Example (3), the APPRECIATION of family harmony is modelled as a pattern of the JUDGEMENT of filial piety and parental responsibility, which is in turn a pattern of the inscribed emotion (i.e. happiness).
As was indicated in Section 2.4.3 of Chapter 2, the current curriculum standards for each subject, including those for EFL education, attaches great importance to emotion and attitude education. It may be justified to contend that the positive APPRECIATION of value encoded in the cartoon complies with the goal of the cultivation of high moral standards that is stipulated in the *Curriculum Standards for English*. It can be inferred from the analysis above that attitudinal meanings in the cartoon of Example (3) include both inscribed AFFECT and invoked APPRECIATION, and the former is more committed than the latter in that it is overtly inscribed rather than indirectly invoked. Through the direct inscriptions as well as implicit invocations, a ‘saturating’ prosody (Martin and White 2005: 19) of positive attitudinal meanings that spreads across the image is established.

As for the verbiage-image relation in Example (3), the verbal texts inserted into the visual image via labeling (for detailed discussion on labeling see Section 4.3.1) introduce the teaching content, i.e. the formal and informal forms of address for family members (i.e. *father (dad), mother (mom), grandfather (grandpa) and grandmother (grandma)*). These address forms are attached as labels for each of the four characters that surround the little girl, who is supposed to use those address forms. The two other words, man and woman, are attached to the father and mother, who are the representatives of typical adult male and female human beings. The verbiage is not attitudinally committed, whereas the image is inscribed with the
emotion of happiness and aims to invoke in the viewer a positive APPRECIATION of family harmony. The whole cartoon integrates the family members into a harmonious whole by putting them in the context of outdoor activity. I attempt to account for the significance of this integrativeness in the invocation of positive APPRECIATION by contrasting it with an example from another series of PEP primary EFL textbooks. Example (4) is extracted from Unit 2 in the Super Kids Students’ Book 4, which illustrates four family members in four separate cartoons.

(4)

[Super Kids Students’ Book for Year 4 2003: 7-8]

In Example (4), the four cartoon characters are independent of each other and spread across two pages. There is no connection between these represented participants, except for the captions underneath the cartoons that indicate their family membership. The four represented participants are depicted in four different contexts: dad is drinking a cup of tea/coffee, relaxing on the sofa; mom is carrying her handbag; sister is holding a lollipop; and brother is riding on a skateboard. In other words, the family members are not represented as a group that is involved in the same activity, but as individuals scattered in various situations. It might be argued that some degree of happiness is inscribed in the two cartoons describing sister and brother. However, this kind of happiness is more of childhood’s joyfulness than family happiness, in that the family members are not presented as a harmonious whole. Neither moral JUDGEMENT nor aesthetic APPRECIATION is invoked in these cartoons. It may be contended that from the perspective of the emotion and attitude education that is highlighted in the current Curriculum Standards for English in China, Example
(3) is more pedagogic than Example (4) due to the fact that along with affectual inscriptions, Example (3) is also attitudinally committed to some degree in terms of APPRECIATION.

### 5.3.1.4 Identification or wishful thinking

Analysis in Section 5.3.1.1 above demonstrates that a positive ATTITUDE (i.e. inscribed AFFECT) towards English-learning, which the editor hopes the textbook readers also assume in the beginning of their foreign language study. It may give rise to the question of possible JUDGEMENT invocations in the cartoons for the primary and junior secondary EFL textbooks. These cartoons represent reality to the viewer in a non-realistic way. The issue of credibility of realistic and non-realistic styles of images will be examined later in Chapter 6; at this stage the focus remains on evaluative meanings, to be specific, the sort of JUDGEMENT inscribed, if any, in the cartoons.

It can be inferred from Table 5-1 above that cartoons are the predominant type of image in the primary and junior secondary EFL textbooks (occupying 97.9% and 89.6% respectively), whereas the standard 35mm photography that complies with what is seen with the naked eye is far less frequently adopted (accounting for 0.2% and 9.3%). Why does the editor, or the image producer, choose to represent the world to the primary and junior secondary students in such a non-realistic style? It may be argued that this is due to the presence of the image producer’s JUDGEMENT of the viewer’s capacity of handling multimodal texts. When designing the cartoon characters and constructing the visual display, the image producer’s conscious decision about manipulating the visual arrangements reflects a certain attitudinal orientation. To be specific, the image producer’s JUDGEMENT concerning the playful character of primary and junior secondary school students and the assumption that these students need fun as one of the motivations to learn are inscribed in the cartoons. Perhaps this could also be interpreted as the manifestation of the image producer’s intention of identifying himself/herself with the students so as to arouse their interest in the images and in language learning. Far less verbiage is provided in the primary
EFL textbooks than the secondary ones, which again echoes the aforementioned JUDGEMENT of primary school students. The following chapter will address the issue of the choice of visual style in relation to the nature of pedagogic discourse.

5.3.2 Portrait and photograph

Portraits and photographs are frequently adopted means of visual representation in the senior secondary EFL textbooks in comparison with their primary and junior secondary ones, accounting for 16.5% and 53.8% respectively as was demonstrated in Table 5-1 above. As can be seen from the analyses below, the verbiage occupies a much higher proportion in the senior secondary textbooks than in their primary and junior secondary counterparts. The photographs in our data can be further divided into two sub-categories, i.e. those close-up shots that are similar to the portraits usually for depicting famous or significant people, and those photographs describing a specific moment of events or activities. The present section will cover both sub-categories, and the portraits whose function is similar to the first type of the photographs will be discussed together with those photographs.

5.3.2.1 Co-instantiation and putative reading

I begin the discussion with serious portraits and the close-up photographs of a single person. Example (5) below provides a good example, which is taken from Unit 5 Nelson Mandela – a modern hero in New Senior English for China Student’s Book 1 (2004: 33-34).

It is a warming-up exercise entitled Pre-reading that precedes a passage for reading comprehension. It verbally and visually presents the information about six important people, which includes six passages illustrated by three portraits (i.e. the upper three images) and three photographs (i.e. the lower three images) respectively. The caption underneath each image indicates the name, nationality, and years of birth and death of the represented participant. The readers, i.e. the senior secondary students, are required to work in pairs and decide upon who are great among these six people, and what makes the difference between great men and important men.
According to the *New Senior English for China Teacher’s Book 1* (2006: 86), five of them (i.e. William Tyndale, Norman Bethune, Sun Yat-sen, Gandhi and Nelson Mandela) are regarded as great people due to the assumption that they have ‘gone through struggles and difficulties for their noble cause’; while Neil Armstrong is not considered to be great because he ‘worked hard to be the first man on the moon but he did not sacrifice anything to do this so he should not be viewed as a great man’. Leaving aside the question of whether the opinion given in the teacher’s book is justifiable or the reasons provided are supportable, I focus on how the verbal texts and visual images in Example (5) work in tandem to attitudinally hint at the intended or putative reading.

(5)

What follows is the verbiage of Example (5), which is numbered by paragraph

*[New Senior English for China Student’s Book 1 2004: 33-34]*
for the ease of analysis. The systems of JUDGEMENT and APPRECIATION will be
drawn upon to analyze the attitudinal meanings encoded.

① William Tyndale (1494-1536), Britain
He wrote the Bible in English so all could read it. He died for his ideas but his
work was later used in the Bible.
② Norman Bethune (1890-1939), Canada
He fought against the German Nazis and Japanese invaders during World War II.
He worked in China as a doctor and saved many Chinese soldiers.
③ Sun Yat-sen (1866-1925), China
He founded the first Republic in China in 1911 after many years’ fighting. He
strongly believed in the three principles: nationalism; people’s rights; people’s
livelihood.
④ Gandhi (1869-1948), India
He gave up a rich life for his ideas and fought for his country to be free from the
UK in a peaceful way.
⑤ Nelson Mandela (1918-), South Africa
He fought for the black people and was in prison for thirty years. He helped the
black people to get the same rights as white people.
⑥ Neil Armstrong (1930-), USA
He landed on the moon in July 1969.

As was reviewed in Section 3.3.1 of Chapter 3, JUDGEMENT is concerned with the
ATTITUDE towards people’s behaviours, whether we admire or criticize them, or praise
or condemn them. Those judgements dealing with ‘normality’ (how unusual
someone is), ‘capacity’ (how capable someone is) and ‘tenacity’ (how resolute
someone is) are grouped under the heading of social esteem, whereas those
judgements concerning ‘veracity’ (how truthful someone is) and ‘propriety’ (how
ethical someone is) are studied under the category of social sanction. APPRECIATION
on the other hand deals with the evaluations of inanimate phenomena and processes,
which encompass ‘reaction’ (to the impact or quality of the thing that is under
evaluation), ‘composition’ (of its complexity or conformity to formal organizational
principles) and ‘valuation’ (based on its social value) (Martin and White 2005: 52-56).
The attitudinal meanings encoded in the verbal texts are analyzed in Table 5-3 below.
Table 5-3 ATTITUDE analysis of the verbal texts in Example (5)

<table>
<thead>
<tr>
<th>Appraising items</th>
<th>Appraiser</th>
<th>Appraised</th>
<th>ATTITUDE</th>
</tr>
</thead>
<tbody>
<tr>
<td>① all could read</td>
<td>Textbook editor</td>
<td>William Tyndale’s work</td>
<td>APPRECIATION</td>
</tr>
<tr>
<td>② saved (many)</td>
<td>Textbook editor</td>
<td>Norman Bethune</td>
<td>JUDGEMENT</td>
</tr>
<tr>
<td>③ strongly believed</td>
<td>Textbook editor</td>
<td>Sun Yat-sen</td>
<td>JUDGEMENT</td>
</tr>
<tr>
<td>④ free</td>
<td>Textbook editor</td>
<td>his (Gandhi’s) country</td>
<td>APPRECIATION</td>
</tr>
<tr>
<td>⑤ peaceful</td>
<td>Textbook editor</td>
<td>Gandhi’s way of liberation</td>
<td>APPRECIATION</td>
</tr>
<tr>
<td>⑤ helped (get the same rights)</td>
<td>Textbook editor</td>
<td>Nelson Mandela</td>
<td>JUDGEMENT</td>
</tr>
</tbody>
</table>

It can be inferred from Table 5-3 that in each of the first five verbal texts in Example (5) the appraiser is the textbook editor, and what is being appraised is the represented participant portrayed in each of the five image or what he has done. As was mentioned above, these five people are considered to be great in the putative reading. As Martin (2000a: 155) comments, APPRAISAL systems can be directly construed in the text via attitudinal lexis, or covertly implicated through the selection of ideational meanings that redound with attitudinal meanings. Positive JUDGEMENT and APPRECIATION are encoded in the first five verbal texts of Example (5). For example, peaceful and free in ④ inscribe APPRECIATION of the manner Gandhi advocated in fighting for the freedom of his nation after the emancipation. Hood and Martin’s (2007) further exploration of the GRADUATION system shows that the scaling of non-attitudinal experiential meanings can invoke ATTITUDE, or more exactly, ‘flag’ ATTITUDE. For instance, all in all could read it in ① above, flags a positive APPRECIATION of the valuation of William Tyndale’s enduring work through ‘quantifying’ a process (i.e. could read it). All quantifies the process as ‘extent’, or to be more precise, as ‘scope in space’, which belongs to the category of FORCE in the GRADUATION system. Another example is for thirty years in ⑤, which quantifies the process (i.e. was in prison) in terms of ‘scope in time’. Admittedly as Hood and Martin emphasize, one needs to read the prosody to nail down the value assigned to the attitudinal meanings. ‘Intensification’ is another category of FORCE along with quantification. It can be infused into the process itself (Hood and Martin 2007: 748).
Cases in point are saved in ②, believe in ③, and helped in ⑤. They intensify the manner as ‘vigour’, thus flagging the positive JUDGEMENT of Norman Bethune’s capacity for surgery, Sun Yat-sen’s tenacity of faithfulness and Nelson Mandela’s propriety of generosity. The intensification infused in a process can be further graded attitudinally with a circumstance of manner, such as strongly believe in ③. Moreover, intensification can also work together with quantification, as in saved many Chinese soldiers in ② and after many years fighting in ③, where many quantifies the entity soldiers and years respectively.

The verbal text concerning Neil Armstrong is exceptional. Evaluative meanings are not inscribed in the exceptionally short text He landed on the moon in July 1969; nor does it employ any options from the GRADUATION system to flag attitudinal meanings. There are no comments for evaluating the moon-landing action or its significance in human history, though it might otherwise be considered as quite an achievement and a variety of linguistic resources could have been drawn upon to inscribe or flag ATTITUDE.

I now look at the images to see how verbal and visual meaning-making resources work in tandem in the construal of the putative reading. The six visual images in Example (5) include two colour portraits (those depicting William Tyndale and Norman Bethune), one black-and-white portrait (the one that depicts Sun Yat-sen) and three chromophotographs (those describing Gandhi, Nelson Mandela and Neil Armstrong). The first five images are similar in the sense that they all take the close-up shots, ranging from the extreme close-ups that mainly show the head (i.e. those of Gandhi and Nelson Mandela), to the big close-ups that display the head and the shoulders (i.e. those of William Tyndale, Norman Bethune and Sun Yat-sen). These five people who are considered to be great men are represented at the close personal distance or even the intimate distance (Hall 1964, cited in Kress and van Leeuwen 2006: 125) from the viewer. A human face is commonly accepted as a major means of recognition that best distinguishes a person. The five images under discussion bring the five great men within arm’s length of the viewers, and thus their facial features are represented in a clear, easily recognized way. The background
patterns of these five images are simple or even monochrome (e.g. those of Sun Yat-sen and Nelson Mandela), and hence the foregrounded represented participants are brought into focus. They are portrayed as typically dressed, with familiar dispositions that add to the ease of recognition. All the efforts mentioned above aim at making these important people’s identities as distinct as possible, with the implication that they are worth remembering. As was indicated above, these five images are mainly serious with no inscribed AFFECT. However, the reverence for these people who are portrayed as great is supposed to be invoked in the putative or intended viewer. Cartoons may also function to make a face identifiable, as often occurs in the personal or political satire that can be found in magazines and newspapers. Nevertheless, cartoons of this kind more often than not impress the viewer by exaggerating or even distorting some facial features of the represented participants, making them difficult to arouse the feelings of profound awe or respect in the viewer. Realistic photographs, on the other hand, do not always involve veneration or admiration. In photojournalism, photographs that show celebrities consciously or unconsciously making unrespectable facial expressions or gestures can often be found. As far as the first five portraits and photographs of Example (5) is concerned, it is the style of the portraiture or photography that conveys a seriousness and solemnity, which is conducive to the invocation of positive JUDGEMENT. In this multimodal text, the images and verbiage work in concert to construe a positive reading position for the reader.

I now turn to the treatment given to the photograph of Neil Armstrong, which differentiates it from the five great people images. In terms of ‘size of frame’, this photograph is a medium close shot, showing Neil Armstrong from the waist up and at the far personal distance. This makes his face quite fuzzy in such a small image. In this photograph Neil Armstrong is dressed in a space suit that indicates his identity as an astronaut. Nevertheless, it is hard to distinguish Neil Armstrong from any other astronaut who may be dressed similarly, which would make it difficult to recognize him next time in another visual image about him. Together with the less value-laden verbal text as analyzed above, this multimodal text represents Neil
Armstrong as an important astronaut, but it does not distinguish him from other astronauts as an influential individual in human history.

Based on the analysis above it can be inferred that the images and verbiage in Example (5) cooperate with one another in the construal of a putative readership. Consequently, the relation between language and images in Example (5) is ‘co-instantiation’ (Martin 2008b). To be specific, positive JUDGEMENT of the people’s great deeds and positive APPRECIATION of their noble causes are inscribed in the verbal texts, which are further reinforced by positive JUDGEMENT invocations through the corresponding portraits or photographs. The meanings made with one semiotic mode are thus ‘multiplied’ (Lemke 1998a: 92) by those that are made with the other. The verbal and visual APPRAISAL resources work in tandem throughout the whole multimodal text, contributing to the construction of the evaluative stance of the appraiser (in this case the textbook editor). This stance defines the kind of community set up around shared values (Martin and Rose 2007a: 59), thus aligning and directing the reader rhetorically into the intended reading position.

5.3.2.2 Co-instantiation in JUDGEMENT invocation

In this section I shall examine the ATTITUDE encoded in the multimodal texts with photographs that capture a specific moment of human activities. The example for analysis is an excerpt from a historical recount entitled A Night the Earth Didn’t Sleep (see Example (6) below), which is one of the reading comprehension materials in Unit 4 Earthquakes in New Senior English for China Student’s Book 1 (2004: 26). What is recorded in this historical recount is the well-known big earthquake that happened in Tangshan in 1976. Several monomodal paragraphs precede the excerpted multimodal text, describing some strange signs before the earthquake, the horrible scenes when the earthquake happened, as well as its disastrous consequences. The verbal text in the multimodal excerpt is the final paragraph concerning the rescue and recovery work, which is accompanied by a photograph that snaps one of the rescue scenes.
The focus of the analysis here is still on the attitudinal meanings, and I start with the language as follows.

All hope was not lost. The army sent 150,000 soldiers to Tangshan to help the rescue workers. Hundreds of thousands of people were helped. The army organized teams to dig out those who were trapped and to bury the dead. To the north of the city, most of the 10,000 miners were rescued from the coal mines. Workers built shelters for survivors whose homes had been destroyed. Fresh water was taken to the city by train, truck and plane. Slowly, the city began to breathe again.

The opening clause *all hope was not lost* functions as the hyper-Theme of the paragraph. On the one hand, it serves as a turning point in the whole essay, shifting the account from the description of the huge damage caused by the earthquake to the timely rescue work that brought hope to the earthquake-stricken people. On the other, it functions to colour a larger segment of the text ‘as a token of evoked APPRAISAL of some kind’ (Page 2003: 216). *All hope was not lost* itself is inscribed with the positive AFFECT, while at the same time it invokes APPRECIATION of reaction that is concerned with the quality of the situation. This in turn elicits a positive prosody of JUDGEMENT that deals with what the army and rescue workers do to help with the salvation. The ATTITUDE analysis of the verbal text in Example (6) is demonstrated in Table 5-4 below, in which the sub-categories of JUDGEMENT and APPRECIATION are identified. Following Martin and White (2005: 71-76), the ideational tokens/invocations are marked with the notation ‘t’ to indicate invoked
ATTITUDE. In addition, the abbreviations below are adopted in the ensuing ATTITUDE analyses.

+ ‘positive ATTITUDE’
- ‘negative ATTITUDE’
des ‘AFFECT: desire’
hap ‘AFFECT: un/happiness’
sec ‘AFFECT: in/security’
sat ‘AFFECT: dis/satisfaction’
norm ‘JUDGEMENT: normality’
cap ‘JUDGEMENT: capacity’
ten ‘JUDGEMENT: tenacity’
ver ‘JUDGEMENT: veracity’
prop ‘JUDGEMENT: propriety’
react ‘APPRECIATION: reaction’
comp ‘APPRECIATION: composition’
val ‘APPRECIATION: valuation’

In this historical recount it is the author who evaluates people and things and tells the reader how to feel, and therefore the author is the source of the APPRAISAL, i.e. the appraiser.

<table>
<thead>
<tr>
<th>Appraising items</th>
<th>AFFECT</th>
<th>JUDGEMENT</th>
<th>APPRECIATION</th>
<th>Appraised</th>
</tr>
</thead>
<tbody>
<tr>
<td>all hope was not lost</td>
<td>des</td>
<td>t, +react</td>
<td></td>
<td>the situation</td>
</tr>
<tr>
<td>To help the rescue workers</td>
<td></td>
<td>t, +prop</td>
<td></td>
<td>the army</td>
</tr>
<tr>
<td>hundreds of thousands of people were helped</td>
<td></td>
<td>t, +cap</td>
<td></td>
<td>the army</td>
</tr>
<tr>
<td>most of the 10,000 miners were rescued</td>
<td></td>
<td>t, +cap</td>
<td></td>
<td>the army</td>
</tr>
<tr>
<td>built shelters for survivors</td>
<td></td>
<td>T, +prop</td>
<td></td>
<td>(the rescue)</td>
</tr>
<tr>
<td>began to breathe again</td>
<td></td>
<td>t, +comp</td>
<td></td>
<td>workers</td>
</tr>
</tbody>
</table>

The analysis above shows that, after the positive APPRECIATION in the opening clause, the evaluation of the army and the rescue workers’ actions is implied through the invocations of JUDGEMENT, which spread across the whole verbal text and explain...
where that hope comes from. Some options from the category of quantification under force are employed to invoke positive attitude. For instance, 150,000, hundreds of thousands of, and most of the 10,000 quantify the entities soldiers (who were sent to help the rescue workers), people (who were helped), and miners (who were rescued). This reinforces the efforts made to rescue the victims and the satisfactory outcome, hence flagging the positive judgement of the propriety and capacity of the soldiers and rescue workers. The prosody of the implied judgement is further confirmed by the concluding positive appreciation of the city that regains its balance after the rescue work. The prosodic realization in the verbal text of Example (6) is ‘domination’ (Martin and White 2005: 20), in which the hyper-Theme all hope was not lost colours and thus tokenizes the attitudinal meanings in the rest of the text. Overall, the verbal text in Example (6) is highly attitudinally committed through the use of both inscribed and invoked attitude.

The photograph that illustrates the verbal text captures the scene in which the army and rescue workers are going all out with basic instruments to dig out the people were trapped in the rubble. In terms of social distance, the photograph is taken at the public distance (Hall 1966: 110-120, cited in Kress and van Leeuwen 2006: 124) from which the viewer can see at least four or five figures with their surrounding environment. Although it is difficult to identify the represented participants individually, the viewer may recognize from the army uniforms that some of them are soldiers. What is implied in this treatment is that there are hundreds of thousands of rescue teams like the one snapped in this photograph, and thus it may be impossible to remember their faces or names. That accords with the verbal text in which all the positive judgement is ascribed to the collective of the army and rescue workers rather than any individual. As for the point of view, the photograph adopts an eylevel, frontal angle, which implies a high level of involvement and presents the represented participants as part of the viewer’s world. The back view puts the represented participants in an easy-to-attack position, and thus the army here is presented as a dedicated, selfless and diligent team rather than a confrontational or aggressive armed force. Finally from the perspective of focalization, the back view
allows the viewer to have a viewer perspective and a character focalized perspective simultaneously (See Section 3.4.2 for review of the conception of focalization). Visually positioned behind the rescue workers and soldiers, the viewer contemplates the disastrous scene along with the rescue team. In other words, the viewer shares with the focalizing represented participants in observing the horrifying aftermath and facing the strenuous and urgent task. Overall, in this photograph the represented participants are portrayed as an altruistic group of rescuers who devote themselves wholeheartedly to the urgent operation at the time of disaster and in the harsh environment. They are part of a far larger rescue team and remain unknown to the public. It can thus be argued that a positive JUDGEMENT of social esteem is invoked in the viewer, or at least in the construed intended viewer, because of the way they are positioned in relation to the photograph. This attitudinal invocation commits a positive evaluative meaning concerning the represented participants.

Taking into account both the verbal text and visual image in Example (6), it can be found that a strong attitudinally cohesive bond is construed between these two semiotic modes. The positive APPRECIATION at the beginning of the verbal text provides colouration and confirmation for the ensuing invoked JUDGEMENT encoded in the selection of ideational meanings. The image on the other hand visually illustrates those ideational meanings while reinforcing the positive evaluation in the verbal text through invoking JUDGEMENT of social esteem in the viewer. The positive attitudinal meanings that are co-instantiated in the verbiage and image develop prosodically throughout the multimodal text, and peak at another positive APPRECIATION at the end of the verbal text that appraises the achievements of the rescue work and confirms the implicit evaluation established by the preceding verbiage and image. This may serve as one way of interpreting how multimodal resources amplify one another (Martin 2004: 298) or in Lemke’s (1998a) term, how the ‘multiplying meaning’ happens.

5.4 Co-instantiation and complementarity in encoding APPRECIATION

The present section will be devoted to the analysis of verbiage-image relations in
appraising the multimodal texts that describe places and civic landscape. I shall position the analysis of language and visual displays within the framework of the APPRAISAL system. As was stated in Chapter 3, among the three categories in the system of ATTITUDE, the dimension of APPRECIATION is an effective tool for analyzing the evaluation of inanimate things, including man-made objects and natural phenomena, concrete entities and abstract concepts.

5.4.1 Invoked APPRECIATION of reaction

Example (7) is extracted from Unit 6 The story of rain in PEP Primary English Students’ Book I for Year 6 (2003: 70). The image in Example (7) chiefly depicts natural scenery. It is a cartoon with such features as the exaggerated size of the eyes of birds, the over-simplified depiction of the cloud patterns, the ripples on the river, and the texture of mountains and trees as well. This image adopts a ‘sensory coding orientation’ (Kress and van Leeuwen 2006: 165; see Section 6.3.2.2. in Chapter 6 for further discussion). One of the most important features of the ‘sensory coding orientation’ is that it is assumed to activate some feelings (e.g. excitement, horror, relief, etc.). In the case of Example (7), the emotions to be stirred up in the viewer are evidently pleasure and excitement. Both the landscape and the creatures are portrayed in the way that it is pleasant to the eye, and the brilliant colouration adds onto this pleasure principle. The verbiage in this multimodal text comes from the character voice outside of the image of landscape. It is represented as the question uttered by the character Wu Yīfan and enclosed in a dialogue balloon or speech balloon. The function of the dialogue balloon here is to instruct the reader to identify what they see in the landscape picture and to put them in English. It may be argued that the small image of Wu Yīfan with the dialogue balloon is an efficient indicator that encourages more reader’s engagement with the landscape picture, and this engagement includes the positive reactions that the picture is supposed to invoke. Overall, the cartoon in Example (7) invokes the response of positive APPRECIATION in the viewer, to be exact, the positive reaction to the arresting, interesting impact of the image on the viewer.
I now turn to the verbiage of Example (7) to outline its interpersonal import. There is no attitudinal meanings encoded in the WH-interrogative clause *What can you see in the picture*. Without using any explicitly evaluative lexis or selected ideational tokens to ‘invite’ or ‘provoke’ an attitudinal response in the reader, the verbiage is hypo-committed attitudinally. On the other hand, as the visual analysis above has demonstrated, the cartoon image is attitudinally committed, which is evident in the pleasing style of the landscape image as well as in the small interactive image of the character *Wu Yifan*. Therefore, it is justified to say that the image and verbiage in Example (7) are complementary in terms of ATTITUDE, since the image is attitudinally committed through the use of invoked APPRECIATION of reaction, whereas in the verbiage there is hardly any attitudinal commitment.

(7)

![Image of a cartoon character asking what can you see in the picture.](PEP Primary English Students' Book I for Year 6 2003: 70)

5.4.2 Co-instantiation in encoding APPRECIATION

The multimodal text I draw on to exemplify the verbiage-image co-instantiation of APPRECIATION is a reading exercise in the teaching section *Using Language* from Unit 2 *The United Kingdom* in New Senior English for China Student’s Book 5 (2004: 13-14). It is composed of a personal recount entitled *SIGHTSEEING IN LONDON,*
narrating the protagonist Zhang Pingyu’s three-day tour in London, and a list-making exercise that is based on her tour and the comments on each place she has visited. The Comments part that the reader is required to fill in the table involves the protagonist’s evaluation of the places, which is encoded in the language and enhanced by the images. For the ease of analysis I list the verbal texts and their corresponding visual displays side by side in Example (8) below. For those who are interested in the exact layout, please see Appendix II for the whole extract.

(8)

Worried about the time available, Zhang Pingyu had made a list of the sites she wanted to see in London. Her first delight was going to the Tower. It was built long ago by the Norman invaders of AD 1066. Fancy! This solid, stone, square tower had remained standing for one thousand years. Although the buildings had expanded around it, it remained part of a royal palace and prison combined. To her great surprise, Zhang Pingyu found the Queen’s jewels guarded by special royal soldiers who, on special occasions, still wore the four-hundred-year-old uniform of the time of Queen Elizabeth I.

There followed St Paul’s Cathedral built after the terrible fire of London in 1666. It looked splendid when first built! Westminster Abbey, too, was very interesting. It contained statues in memory of dead poets and writers, such as Shakespeare. Then just as she came out of the abbey, Pingyu heard the famous sound of the clock, Big Ben, ringing out the hour. She finished the day by looking at the outside of Buckingham Palace, the Queen’s house in London. Oh, she had so much to tell her friends!

The second day the girl visited Greenwich and saw its old ships and famous clock that sets the world time. What interested her most was the longitude line. It is an imaginary
line dividing the eastern and western halves of the world and is very useful for navigation. It passes through Greenwich, so Pingyu had a photo taken standing on either side of the line.

The last day she visited Karl Marx’s statue in Highgate Cemetery. It seemed strange that the man who had developed communism should have lived and died in London. Not only that, but he had worked in the famous reading room of the Library of the British Museum. Sadly the library had moved from its original place into another building and the old reading room was gone. But she was thrilled by so many wonderful treasures from different cultures displayed in the museum. When she saw many visitors enjoying looking at the beautiful old Chinese pots and other objects on show, she felt very proud of her country.

The next day Pingyu was leaving London for Windsor Castle. “Perhaps I will see the Queen?” she wondered as she fell asleep.

The verbiage of Example (8) is value-loaded and the ATTITUDE analysis is presented in Table 5-5 below. Generally speaking, the author of the passage is considered to be the ultimate source of APPRAISAL. Nevertheless, when the attitudinal meanings are specifically ascribed to certain participant, I shall analyze the participant as the appraiser.

<table>
<thead>
<tr>
<th>Appraising items</th>
<th>Appraiser</th>
<th>AFFECT</th>
<th>APPRECIATION</th>
<th>Appraised</th>
</tr>
</thead>
<tbody>
<tr>
<td>Worried</td>
<td>Zhang Pingyu</td>
<td>- sec</td>
<td></td>
<td>time</td>
</tr>
<tr>
<td>Wanted</td>
<td>Zhang Pingyu</td>
<td>+des</td>
<td></td>
<td>sites</td>
</tr>
<tr>
<td>Delight</td>
<td>Zhang Pingyu</td>
<td>+hap</td>
<td></td>
<td>going to the Tower</td>
</tr>
<tr>
<td>Fancy</td>
<td>Zhang Pingyu</td>
<td>+reac</td>
<td></td>
<td>the Tower</td>
</tr>
<tr>
<td>Solid</td>
<td>Zhang Pingyu</td>
<td>+comp</td>
<td></td>
<td>the Tower</td>
</tr>
<tr>
<td>standing for one thousand years great surprise</td>
<td>Zhang Pingyu</td>
<td>neutral</td>
<td>t, +comp</td>
<td>royal soldiers’</td>
</tr>
</tbody>
</table>

[New Senior English for China Student’s Book 5 2004: 13-14]
As is shown in Table 5-5 above, the attitudinal meanings are either overtly inscribed or implicitly encoded through the selection of ideational tokens. Among the twelve overt inscriptions of APPRECIATION (i.e. fancy, solid, splendid, interesting, famous, famous, what interested her most, very useful, strange, famous, wonderful, and beautiful) eleven of them are positive evaluations, which account for nearly 92% of the total APPRECIATION. These include the reactions to the captivating nature of the places of interest, the evaluations of the durable and substantial quality of the Tower of London, and the recognition of the high value of the longitude line in navigation. The one and only lexical item with negative notation is the word strange, which indicates a sense of counter-expectancy during the protagonist’s visit to Highgate Cemetery. Along with the explicit inscriptions of APPRECIATION, there are selections of ideational meanings (i.e. had so much to tell, sets the world time, and dividing the
eastern and western halves of the world) to invoke evaluations of the scenic spots that are described. Moreover, quantification is adopted to invoke ATTITUDE. For example, in standing for one thousand years the entity years is quantified as the circumstance of manner standing, which aims at flagging a positive APPRECIATION of the long history and solid composition of the Tower of London. For instance, the expression of so much to tell implies the fruitfulness of Zhang Pingyu’s first day’s tour. Standing for one thousand years provides convincing evidence for the evaluation of the solidness of the Tower of London, while sets the world time and dividing the eastern and western halves of the world confirm the assessment of the great value of the clock and longitude line in Greenwich. In a word, the selections of ideational tokens reinforce the explicit attitudinal lexis, while the overt attitudinal inscriptions confirm the implicit invocations by explicitly telling the reader how to feel.

In terms of the nine attitudinal lexical items that are inscribed with affectual meanings (i.e. worried, wanted, delight, great surprise, sadly, thrilled, enjoying, very proud, and wondered), six of them belong to positive AFFECT. They encode Zhang Pingyu’s as well as other visitors’ great delight during the tour and their strong interest in the sightseeing. Great surprise is arguably a neutral feeling that can be approached ‘as a separate type of AFFECT’ (Bednarek 2008: 161). Based on the studies on corpus data, Bednarek (2008: 161-165) articulates that surprise is not construed culturally as a negative or positive emotion. Rather, it is ‘relatively neutral’, and the positivity or negativity of this emotion ‘can be contextually implied with the help of other evaluations preceding or following the emotion term’. In Example (8), the protagonist felt surprised at the fact that the present-day royal soldiers still wore the four-hundred-year-old uniform on special occasions, which highlights what is unusual or interesting and thus implies a positive reaction. Retrospectively, the positive AFFECT delight at the beginning of the paragraph confirms the positive attitudinal orientation of the term great surprise. The remaining two negative inscriptions are designed to invoke positive APPRECIATION. To be specific, at the beginning of the recount Zhang Pingyu was worried about the limit of time so she had made a list to organize the three days’ tour before the
sightseeing, which foretells the value and fruitfulness of the tour to be described later in the recount. Towards the end of her London tour, Zhang Pingyu felt sad because the Library of the British Museum had moved and the old reading room where Karl Marx had worked was gone. Nevertheless, the adversative conjunction but and the positively inscribed affectual lexis thrilled that are immediately after the negative emotion indicates that the tour is still highly rewarding. Overall, through the overt demonstration of how the protagonist feels, the inscribed AFFECT first initiates and then reinforces the prosody of the positive ATTITUDE throughout the verbal text.

Owing to the fact that there are no human participants depicted in the four accompanying photographs of selected buildings or statues, the visual analysis here mainly focuses on angle, or point of view. Along the vertical axe, all the five photographs are taken from a low angle, that is, the photographer and the viewer look at those buildings depicted in the photographs from below. As Kress and van Leeuwen (2006: 140) point out, in this kind of image the represented participants have power over the interactive participants, which makes the former look awesome. The low angle in the five photographs endows the places of interest with symbolic power and authority, though a matter of degree of the power is involved. The St Paul’s Cathedral, Westminster Abbey and Royal Observatory in Greenwich are presented in the photographs as steep tower above the viewer, whereas in the photographs of Royal Observatory in Greenwich and the Karl Marx’s statue they show relatively gradual incline. In terms of the horizontal angle, the scenes are mainly photographed from an oblique point of view and thus indicate a sense of detachment from the viewer, which enhances the meaning of awe and magnificence. The Karl Marx’s statue is represented in a different way from the buildings, in that this photograph is taken from the frontal angle so as to make the statue identifiable to the viewer. To sum up, the interplay of the vertical and horizontal angles encodes the ATTITUDE of reverence towards the civic landscapes in London, which co-articulates the APPRECIATION and AFFECT encoded in the verbal text. In other words, both the verbiage and the images commit positive attitudinal meanings, which together contribute to the positive prosody of evaluation of the landscapes in the multimodal text.
5.4.3 Co-instantiation in jointly-constructed text

I now attempt to examine the distribution of attitudinal meanings in multimodal jointly-constructed text. The example I draw on here is the one that was used in Section 4.3.3.3 (see Example (9) below), which is taken from Unit 4 Earthquakes in the senior secondary EFL textbook *New Senior English for China Student’s Book 1* (2004: 30). In this multimodal jointly-constructed text, the reader is required to participate in the construction of verbiage in light of the accompanying images. The focus of the present discussion will be shifted from the management of heteroglossic space as was analyzed in Chapter 4, to the examination of the verbiage-image relation in encoding attitudinal meanings.

The images of Example (9) is composed of a set of four stamps entitled *The New Tangshan after the Earthquake*, which was issued in 1996 to honour the city twenty years after the Tangshan earthquake. The stamps adopt the commonsense naturalistic coding orientation (Kress and van Leeuwen 2006: 165-166) to depict the civic landscapes of Tangshan after the city reconstruction, portraying *Farmhouses*, *Factories*, a *Street Vista* and the *Harbour* respectively. The scenes are represented from a high angle, putting the viewer at the position of overlooking the whole picture.
On the left of the stamps there is a speech draft to be completed by the reader. As is indicated in the exercise instruction, the reader is required to give a short talk about the new stamps by using the model provided in the draft.

Thank you, Mr Zhang. I am happy to ____________. As you can see, the stamps show ____________. I think these stamps are very important because ____________. I will collect the stamps ____________. It will be my way to honour all the people who lost their lives in the earthquake and ____________. Thank you for _____________.

The attitude encoded in this short speech draft covers all three dimensions, i.e. AFFECT, JUDGEMENT and APPRECIATION, as is shown in Table 5-6 below. The ellipsis notation ‘…’ in the table indicates what is to be filled up by the reader.

<table>
<thead>
<tr>
<th>Appraising items</th>
<th>Appraiser</th>
<th>AFFECT</th>
<th>JUDGEMENT</th>
<th>APPRECIATION</th>
<th>Appraised</th>
</tr>
</thead>
<tbody>
<tr>
<td>Happy</td>
<td>I</td>
<td>+hap</td>
<td></td>
<td></td>
<td>...</td>
</tr>
<tr>
<td>very important</td>
<td>I</td>
<td></td>
<td>+val</td>
<td>stamps</td>
<td></td>
</tr>
<tr>
<td>Honour</td>
<td>I</td>
<td></td>
<td>+prop</td>
<td>people who lost their lives in the earthquake and …</td>
<td></td>
</tr>
</tbody>
</table>

As is indicated in Table 5-6, the target of personal emotions is to be completed by the reader, while the evaluation directed towards the stamps are clearly specified as very important. The reader is supposed to supply another target of the positive judgement indicated by the verb honour, in addition to the victims of the earthquake already provided in the text. It can be argued that what are supposed to be filled in the blanks are closely related to what are depicted in the visual images. In order to gain a better understanding of the verbiage-image relation, I draw upon a sample answer in the corresponding teacher’s book (2006: 70). For the sake of easy analysis, I number each of the blanks in the sample draft.

Thank you, Mr Zhang. I am happy to ① talk about the new stamps. As you can see, the stamps show ② different scenes from our new city. I think these stamps are very important because ③ they will let others around the
country remember what we have done here. I will collect the stamps ④ because when I am old, I can show my grandchildren these stamps and tell them about the earthquake. It will be my way to honour all the people who lost their lives in the earthquake and ⑤ the people who have done their best to rebuild the city. Thank you for ⑥ letting me make this speech.

It can be inferred from the draft above that most of the answers are closely linked to the visual displays in the stamps. For example, ① is concerned with the stamps themselves; ② and ⑤ are related to what are described in the stamps; and ③ and ④ are about the speech maker’s imagination based on the stamps. It is noteworthy that there is no explicit attitudinal lexis in the sample answer. This is because the overall ATTITUDE of the speech has generally been prescribed in the scaffolding of the draft, which is already provided before the reader completes the jointly-constructed text. In other words, what are required to be filled in the blanks are the ideational meanings based on the accompanying visual images of the stamps. Admittedly, certain selections of these ideational tokens may serve as implicit attitudinal invocations. For instance, let others around the country remember what we have done here may invoke positive reaction to what we have done here. I can show my grandchildren these stamps and tell them about the earthquake is assumed to invoke positive evaluation about the stamps. The people who have done their best to rebuild the city may invoke positive JUDGEMENT of the people’s efforts. The reader is thus required to provide ideational tokens for the attitudinal inscriptions in the verbal text by drawing upon the hints from the images. To sum up, it is argued that the verbiage-image relation in this jointly-constructed text is co-instantiation, in that the verbal text establishes a prosody of positive evaluation, whereas the visual images provide the ideational tokens that are indispensable in completing the jointly-constructed text, and at the same time they can also invoke positive ATTITUDE in the viewer. On the one hand, the attitudinal meanings implied in the selection of ideational tokens in the visual images are confirmed by the explicit ATTITUDE in the verbal text, and on the other they further reinforce the overt attitudinal inscriptions in the verbiage.
5.4.4  A note on degrees of commitment in visual communication

As was mentioned in Section 3.2.3 of Chapter 3, the concept of commitment, i.e. the amount of meaning potential that is instantiated in a given text, can be employed to capture how explicit or implicit a certain instance encodes attitudinal meanings. Based on the comparative study of three verbal accounts (one biographical and two auto-biographical), Martin (2008b) points out that direct inscription commits more attitudinal meaning potential than invocation does; while between the sub-categories of invocation, flagging is more attitudinally committed than affording. If this discussion is extended to include visual meaning-making resources that encode attitudinal meanings, it is justified to argue that in our data Examples (1), (2), (3) and (7) above employ depiction techniques that explicitly show the viewer how the represented participants feel (e.g. happy in these cases) and thus overtly tell the viewer about their emotions. There is no verbal commitment in terms of ATTITUDE in these four examples, and it is the visual images that are responsible for the attitudinal commitment. In Examples (5), (6), (8) and (9), language joins images in encoding attitudinal meanings, either through the use of evaluative lexis or via the selections of ideational tokens. These images do not embrace visual patterns that encode AFFECT, and yet the ideational selections of certain visual representations do seem to be intended to invoke an attitudinal response (i.e. JUDGEMENT or APPRECIATION) in the viewer. The indirect invocations in the images of Examples (5), (6), (8) and (9) may not commit as much ATTITUDE as the direct inscriptions in the images of Examples (1), (2), (3) and (7), in that the viewer need to take a crooked path to grasp the attitudinal meanings encoded in the images. Nevertheless, the co-instantiation of both language and images in the former co-articulate each other throughout the unfolding of a multimodal text.

In those cases in which only visual patterns commit attitudinal meanings, the verbiage-image relation is that of complementarity rather than co-instantiation. The visual images complement the corresponding verbal texts in the sense that they are where the attitudinal meanings come from, whereas the verbal texts complement the images by providing the materials that are designed to meet certain teaching goals.
Here I attempt to make some critical comments on those multimodal texts for primary and junior secondary education, in which there is far too little verbiage as compared with the visual images. It is obvious that these texts follow the traditional bottom-up approach that advocates teaching language should be conducted from the smaller units (e.g. words, groups) to the larger ones (e.g. clauses, texts). This approach, as is pointed out by Martin and Rose (2007b), reverses the course of natural language learning, in which linguistic features are experienced and encountered in meaningful contexts as texts rather than memorized as standards or rules. Rather, language learners should be guided to recognize and use the language patterns appropriately in predictable social contexts. Based on the Sydney School genre theory (Martin 1992, 2000b; Martin and Rose 2007a, 2008) that maps cultures as systems of genres from a semiotic perspective (Martin and Rose 2008: 17), the genre-based literacy approach demonstrates that the teaching of reading and writing should start with the global social contexts in which texts are embedded, examining the generic/schematic structures, and then move downwards to the recognition of the linguistic features that are specific to different genres at the levels of semantics, lexicogrammar and phonology (for more discussion concerning the stratification within language system and between semiotic strata, see Section 3.2.2). In Examples (1), (2), (3) and (7) above, there is too little verbiage to construe the verbal context that is indispensable to a solid understanding of the social purposes or functions that the texts aim to achieve. Therefore, it may be justified to argue that while paying regard to the functions of visual images in managing heteroglossic space as well as encoding attitudinal meanings, textbook editors should also acknowledge the importance of verbal context and keep in mind the fact that EFL teaching should be approached from the level of genre, then the discourse semantic patterns, which is followed by the lexicogrammatical realizations.

5.5 The accumulation of institutionalized feelings in EFL education

In the preceding two sections I have examined what sorts of attitudinal meanings are
encoded in the various types of visual displays in the multimodal EFL textbook discourse, and analyzed the ways in which image and verbiage co-articulate or complement each other in construing evaluation. Now I am in a better position to discuss the encoding of attitudinal meanings in the primary and secondary EFL textbooks in relation to the emotion and attitude goal in the curriculum standards, by drawing on the conception that views JUDGEMENT and APPRECIATION as institutionalized feelings (Martin 2000a; Martin and White 2005). Before doing so, I shall first briefly mention what is emphasized in the current curriculum standards as emotion and attitude education.

5.5.1 Emotion and attitude goal in the current curriculum standards

In the educational context in China, the best known aspect of emotion and attitude education might be moral and values education. As has been observed by Zhu (2006), moral and values education in China had for a long time been conducted mainly through ‘a special system for political work’, including Party organization, Youth League organization and Young Pioneer organization, as well as through ‘a special subject for moral education’ in schools. It is the first time in the curriculum standards published by the Ministry of Education of the People’s Republic of China in 2001 that emotion and attitude education, which includes moral and values education, is defined from the angle of ‘the integrity and conformity of curriculum functions’.

Emotion and attitude education is seen as indispensable to subject teaching in the reformed curriculum. The current curriculum standards for each subject have a clear statement on the attitudinal goal of developing students’ active emotion (Zhu 2006). Section 2.4.3 in Chapter 2 has provided examples and explanations concerning the right orientation of emotions, attitudes, and values in the Curriculum Standards for English. The following excerpt is taken from Article I in Part II of Curriculum Standards for Mathematics, which elaborates on the emotion and attitude goals in mathematics education.
Emotion and attitude goals:

- To actively engage in mathematics learning activities; full of curiosity and desire for learning mathematics;
- To have the experience of success in mathematics learning, forge the will to overcome difficulties, and build up confidence;
- To preliminarily understand the close relationship between mathematics and life, as well as the role of mathematics in human’s historical development; to experience mathematical activities that are full of exploration and creation; realize the necessity of preciseness and the certainty of the conclusion;
- To assume the attitude of respecting objectivity, and to form the habit of making critical comments and thinking independently.

(Zhonghuarenmingongheguojiaoyubu 2001b: 7)

It can be inferred from the two curriculum standards that emotion and attitude education in the current curriculum standards in China embraces having a positive attitude, confidence, a desire to learn, and an ability to evaluate people’s behaviours as well as things and phenomena. Language learning is regarded as ‘far more closely related to emotion and attitude education than other school subjects’ (Cheng 2002). As was indicated Chapter 2, the current Curriculum Standards for English adopts a nine-grade system for stipulating the specific goals in EFL teaching and learning at each educational level. In terms of emotion and attitude education, what are assumed to be achieved attitudinally by the students of Grades 2, 5, and 8 (i.e. primary, junior and senior secondary school graduates) are summarized in Table 5-7.
Table 5-7 Emotion and attitude goals for Grades 2, 5, and 8 in *Curriculum Standards for English*
(Zhonghuarenmingongheguo jiaoyubu 2001a: 22)

<table>
<thead>
<tr>
<th>Grade</th>
<th>Description of goal</th>
</tr>
</thead>
</table>
| 2     | 1. Interested in listening to and speaking English, as well as other activities like reciting chants, singing songs, telling stories, and playing games;  
       2. Happy to follow good models; dare to open one’s mouth and speak English; actively engage oneself and take the initiative to consult others. |
| 5     | 1. Clear about the purpose for learning English; aware of the fact that learning English is for communication;  
       2. Interested in learning English and willing to do so, happily engaged in a variety of English-learning activities and practices;  
       3. Confident about mastering English; dare to express oneself in English;  
       4. Able to actively cooperate with others, help one another, and accomplish tasks together;  
       5. Happy in learning English and willing to learn English songs as well as reading materials in English;  
       6. Able to pay regard and understand others’ feelings while communicating in English;  
       7. Able to take the initiative to consult teachers or peer students and ask them for help when encountering problems;  
       8. Happy to explore English and follow good models in everyday life;  
       9. Have a deeper understanding of Chinese culture;  
      10. Happy to learn about foreign culture. |
| 8     | 1. Maintain interest in learning English as well as the desire to learn, and actively engage with those activities that are conducive to improving one’s English;  
       2. Have a proper motivation for learning English; clear about the fact that learning English is for communication and expression;  
       3. Have relatively strong confidence in learning English, dare to communicate with others and express oneself in English;  
       4. Able to overcome difficulties in learning English, and willing to consult others;  
       5. Able to understand and respect others’ feelings while communicating in English;  
       6. Have relatively strong teamwork spirit in learning; willing to share learning materials with others;  
       7. Able to introduce Chinese culture in English;  
       8. Able to learn about and respect foreign culture, as well as demonstrate the spirit of international cooperation. |

As is demonstrated in Table 5-7 above, the emotion and attitude goals for Grade 2 are mainly about inborn emotions (e.g. interested, happy) towards English learning,
along with some indications of tenacity required in the learning process (e.g. *dare to open one’s mouth, take the initiative*). As for Grade 5, while positive emotions like happiness, security and satisfaction (e.g. *interested, happy, willing and confident*) are retained, more requirements are made in terms of capacity (e.g. *clear, aware, and able to actively cooperate... to pay regard and understand...*), tenacity (e.g. *take the initiative*), as well as the evaluation of Chinese and foreign cultures. When it comes to Grade 8, except for Items 1 and 3 that mainly deal with emotions, the goals primarily stated for senior secondary students to achieve centre around capacity (e.g. *clear, able to overcome difficulties... to understand and respect... to learn about and respect...*) and propriety (e.g. *teamwork spirit, willing to share*). Based on the description in the previous sections of the ways in which image and verbiage work in tandem to set up the positive evaluative orientation, I will discuss in the ensuing section on the relationship between the three dimensions within ATTITUDE and its implications for the emotion and attitude education.

### 5.5.2 JUDGEMENT and APPRECIATION as institutionalized AFFECT

As was indicated in Section 3.3.1 in Chapter 3, among the three attitudinal sub-systems AFFECT can be taken as the core system because it is concerned with the embodied feelings we are born with, which is later developed into culturally specific emotional repertoires (Martin 2000a, 2004; Painter 2003). AFFECT is recontextualized, or institutionalized into the realms of shared values in two directions: in relation to the evaluation of human behaviors that need to be controlled under social norms, AFFECT is recontextualized as JUDGEMENT; and with respect to the evaluation of things that need to be valued with reference to their social significance, AFFECT is recontextualized as APPRECIATION. In other words, JUDGEMENT and APPRECIATION can be regarded as institutionalized AFFECT. Based on Halliday’s (1994) concepts of proposition and proposal, Martin (2000a: 147) suggests that JUDGEMENT institutionalizes feelings as proposals (about behaviour), whereas APPRECIATION institutionalizes feelings as proposition (about things). Figure 5-3 below shows diagrammatically these two recontextualizations or institutionalizations.
In developing an ontogenetic view on attitude in language development, Painter (2003) studies the ontogenesis of ATTITUDE system and points out that a child’s initial semiotic system is essentially one for sharing emotion or AFFECT. The institutionalizations of AFFECT, i.e. JUDGEMENT and APPRECIATION, occur ‘only later with new semiotic steps on the child’s part—either the adoption of mother tongue lexical words in lieu of protolinguistic symbols or a “meta” awareness of the sign itself’ (Painter 2003: 189). In other words, the growing capability of symbolizing enables the evaluations in terms of JUDGEMENT and APPRECIATION to emerge along with the direct expressions of emotions.

Figure 5-3 JUDGEMENT and APPRECIATION as institutionalized AFFECT
(Martin and White 2005: 45)

Here I attempt to extend the discussion beyond early pre-school period to include primary and secondary education, which may account for the construal of evaluation in the given educational context. There are a wide range of semiotic resources, both visual and verbal, available to construe evaluative stance in EFL textbooks. As was analyzed above, the images in the textbooks for primary students mainly adopt cartoons to describe animate represented participants with explicit affectual inscriptions. This tendency gradually diminishes and images of serious style (i.e. portraits and photographs) emerge in the textbooks for secondary education. Combined with the ideational tokens that are designed to invoke APPRAISAL as well as
the explicit JUDGEMENT in the verbal texts, the serious portraits and photographs encode JUDGEMENT invocations and colour the whole multimodal text. As for those multimodal texts that depict inanimate buildings and landscapes, visual images step up to invoke APPRECIATION in the viewer where the verbal texts are not responsible for the ATTITUDE-encoding. In those multimodal texts in which verbal texts and visual images are both attitudinally committed, images enhance the verbally inscribed or invoked attitudinal meanings. This co-instantiation contributes to the establishment of the intended evaluative stance. To sum up, in examining the visual and verbal attitudinal inscriptions and invocations in EFL textbooks as the students advance through school years from primary to secondary education, an attitudinal shift from a more personal way of expressing individual emotions to a more institutionalized way of evaluating people’s behaviours, abstract concepts, and concrete objects can be identified. This change of preference both reflects and contributes to the gradual increase of institutionalized feelings that are supposed to be achieved, as is indicated in the current pedagogic context.

5.6 Summary

In sum, the current chapter is concerned with the evaluative orientation construed in the multimodal EFL textbook discourse, with emphasis placed on the verbiage-image relations. To be specific, I have analyzed the attitudinal meanings encoded in various styles of visual images that are widely employed in the textbooks under discussion, i.e. cartoons, portraits and photographs, and their correlated verbal texts. The research findings indicate that the frequently-adopted cartoon style in the primary textbooks is often indicative of positive affectual meanings, and thus the emotion of happiness in English-learning is directly inscribed. Some of the cartoons can also function to invoke APPRECIATION through logogenetic recontextualization. In short, the visual images of cartoons bear much of the responsibility for encoding attitudinal meanings in the textbooks for primary and junior secondary education, with the verbiage-image relation being a complementary one. These attitudinally
under-committed verbal texts are criticized for the adoption of the bottom-up approach in EFL teaching, which disregards the global social context as well as the course of natural language learning.

The trend for explicit affectual inscriptions gradually declines in the textbooks for secondary education. The emerging serious images signal a transition from the playful childhood to the adolescence, when a sense of social responsibility is to be cultivated. The serious portraits and photographs with the implicit JUDGEMENT or APPRECIATION invocations work in concert with the explicit attitudinal inscriptions and the ideational tokens designed to invoke APPRAISAL in the verbal texts. In other words, both language and images commit attitudinal meanings in these multimodal texts, and thus the verbiage-image relation here is co-instantiation. The two semiotic modes co-articulate each other in leading the reader to the putative reading and guiding him/her in the completion of jointly-constructed texts.

In the current curriculum standards, emotion and attitude education has gained unprecedented attention. Taking a multilevel view on the attitudinal meanings inscribed in or invoked by the multimodal EFL textbook discourse, I have attempted to grasp the gradual shift in ATTITUDE -encoding, i.e. from a personal emotion release to an institutionalized way of evaluation, which both reflects and contributes to the orientations of emotion and attitude education that are stipulated in the current curriculum standards.

NOTES

1 The present study on gradability of attitudinal meanings follows Martin and White’ s (2005: 48) view that the high, median, and low degrees do not imply discrete values, but are used to describe the emotions graded along a clined scale.

2 A system network representing the possibilities of the sub-systems coupling with each other could be produced with more visual instances analyzed in future research.
Nelson Mandela’s photograph may be interpreted as slightly impassioned, and yet I consider that it aims at capturing the temperament of Nelson Mandela instead of expressing affectual meaning. Actually whether the affectual meaning encoded in this image is joy or anger or some other feelings is difficult to nail down.
Chapter 6

Contestable Reality: A Multilevel View on Modality in Visual Communication

6.1 Introduction

In this chapter I shall explore the issue of ‘what counts as real’ in multimodal communicative contexts. Social semiotic approach to modality holds that the reliability of a message is socially defined according to the accepted criteria for truth in the intended social or institutional group, which in turn both reflects and constructs social relations (Hodge and Kress 1988; Kress and van Leeuwen 2006; van Leeuwen 2005). Before embarking on the detailed analysis and discussion, I shall first briefly review the linguistic and social semiotic underpinnings that are relevant to the present study, and then introduce several theoretical concepts that are significant in understanding visual modality. The ensuing analysis is a multilevel, comparative investigation into the truth criteria in the primary, junior and senior secondary EFL textbooks, whose purpose is to clarify the connection between modality judgements and social relations in different EFL educational contexts. This will be followed by a discussion on visual style in relation to some major educational theories in the twentieth century. Attention will also be paid to the possibility of choosing different visual styles in constituent genres within a teaching unit as macrogenre.

6.2 Modality in multimodal discourse

The study of modality began with the ‘absolute, context-independent truth of
assertions’ in the philosophy of language before it moved to linguistics (van Leeuwen 2005: 165). The notion of ‘modality’ has been adopted and approached by linguists, logicians, and social semioticians. This section aims at providing a general account of some previous studies on modality in language and other modes of communication that are relevant to the focus of the present research.

6.2.1 Linguistic modality
Traditionally, the term ‘modality’ is generally defined as ‘the manner in which the meaning of a clause is qualified so as to reflect the speakers’ judgement of the likelihood of the proposition it expresses being true’ (Quirk et al. 1985: 219). The studies of modality have for a long time been primarily confined to the discussion of modal auxiliary verbs such as can, could, must, may, might, will, would, etc. In traditional grammar, the meanings of modality are divided into two categories, i.e. intrinsic modality and extrinsic modality. The former covers the meanings of permission, obligation, and volition, whereas the latter includes possibility, ability, necessity, and prediction. Each of the modal auxiliary verbs is believed to have both intrinsic and extrinsic uses (Quirk et al. 1985: 219-221). The distinction between epistemic modality and deontic modality (Coates 1983; Palmer 1990) again reflects these two different uses of the modals. It is argued that most of the modals are used in both senses, and modal verbs can be approached in terms of kinds of modality (e.g. epistemic, deontic and dynamic) and degrees of modality (e.g. possibility and necessity) (Palmer 1990: 8, 36).

As was explained in Chapter 3, the theoretical rationale that underpins the present study owes most to Halliday (1976, 1994). According to Halliday (1994: 88, 356), modality refers to the intermediate degrees of meaning between positive polarity and negative polarity. In English the choices of ‘yes’ and ‘no’ are expressed in the Finite element of the mood structure or Mood Adjunct like no, while the various kinds of interminacy in between are conveyed through modality, which is categorized into modalization and modulation based on whether it is concerned with a proposition or proposal. Modalization is subdivided into probability and usuality to express
different degrees of likelihood (e.g. *possibly, certainly*) and oftenness (e.g. *sometimes, always*). Modulation, on the other hand, consists of the sub-categories of obligation (e.g. *supposed to, required to*) and inclination (e.g. *willing to, determined to*) (Halliday 1994: 88-89, 356-357). The inclusion of modulation into the modality system, as van Leeuwen (2005: 165) comments, has moved the traditional focus on the modality of representations a step further by pointing out the fact that the same modal auxiliary verb (e.g. *may*) can express a degree of social obligation (e.g. *You may go home now*) as well as a degree of representational truth (e.g. *It may rain today*).

Halliday (1994: 355-363) further identifies the three variables in modality: type, orientation, and value. Types of modality include probability, usuality, obligation, and inclination, which represent different ways of construing the semantic space between the positive and negative poles. The second variable ‘orientation’ is used to discuss how each type of modality can be realized. To be specific, there are distinctions between subjective modality and objective modality, and between explicit and implicit variants. It should be noted that the explicitly subjective (e.g. *I think Alice likes it*) and explicitly objective (e.g. *It is likely that Alice likes it*) forms of modality are metaphorical variants, in that they both represent the speaker’s point of view through a projecting clause rather than an adjunct to a proposition. In the explicitly subjective modality, the speaker chooses to highlight his/her own position in assessing the validity of the proposition or the appropriateness of the proposal. On the contrary, in the explicitly objective modality the speaker dresses up his/her angle as objective, hence concealing the source of the attitude being expressed. The third variable is the value that is attached to modality. Take probability for instance. The value can be high (e.g. *It is certainly that Tom knows the truth*) or median (e.g. *It is probably that Tom knows the truth*) or low (e.g. *It is possibly that Tom knows the truth*).

As has been pointed out by Hodge and Kress (1988: 124), one of Halliday’s fundamental contributions to the studies of modality lies in his extension of the use of modality beyond auxiliary verbs to cover all the elements with the same function,
including nouns (e.g. *It is a matter of fact that...*), verbs (e.g. *I doubt that...*), adjectives (e.g. *It is foolish to deny that...*), and other expressions such as *kind of*, *hardly* and *the like*. Furthermore, Halliday suggests that modality is part of the interpersonal component of his functional grammar, rather than placing it in the ideational component as most logicians have assumed. To quote Hodge and Kress,

Halliday’s theory recognizes that modality is a matter of the relation of the participants in a verbal interaction, hence squarely in the domain of the social, and that modal forms are the traces of the activity of speakers acting in a social context. (Hodge and Kress 1988: 124)

Instead of considering modality as expressing the actual objective truth that exists between propositions and the real world, social semioticians examine it in relation to the semiotic resources that are employed to express ‘as how true or as how real a given representation should be taken’ (van Leeuwen 2005: 281; emphasis in the original). The social semiotic approach to modality in visual communication will be explicated below, before developing a multilevel view on modality in texts in different educational contexts.

### 6.2.2 Social semiotic approach to modality

Social semioticians (e.g. Hodge and Kress 1988; Hodge and Tripp 1986; Kress and Hodge 1979; van Leeuwen 1999, 2005) adapt the term ‘modality’ from linguistics to the studies of other semiotic phenomena. In the eyes of social semioticians, modality is used to describe ‘the stance of participants in the semiotic process towards the state and the status of the system of classification of the mimetic plane’ (Hodge and Kress 1988: 122). Therefore, when exploring modality analysts need to examine the semiotic resources that are available to a speaker or writer in negotiating how true a given representation should be taken.

Drawing social semiotic approach to modality, the present research holds that ‘truth’ is the shared truth produced by members of a group or community with the
same value, and it is closely related to ‘affirmation of solidarity’ and ‘assertion of power’ (Hodge and Kress 1988). Modality always involves at least two parties, and the degree of affinity expressed in modality posits the social relationship on a continuum from affirmation (i.e. high affinity/modality) to negation (i.e. weak or zero affinity/modality) (Hodge and Kress 1988: 122-123, 164). In social semiotics, all semiotic processes are treated as social processes, and thus all semiotic phenomena bear signs of modality. To take linguistic phenomena for example, some genres (e.g. encyclopaedia) are regarded as more factual than others (e.g. fiction). What is regarded to be credible to a certain social group may not be reliable in the eyes of another social group (Hodge and Kress 1988: 121; Kress and van Leeuwen 2006: 171). Whoever controls modality can control which version of reality is treated as the valid version, and the accepted representation of reality will serve as ‘the basis of judgement and action’ (Hodge and Kress 1988: 147). In sum, truth can be constantly challenged and tested in every social or semiotic exchange, and modality is a social or interpersonal concept instead of an ideational one.

As was indicated above, for social semioticians modality is a multimodal concept that goes beyond language and is extended to the non-verbal modes of communication. Research on modality in other semiotic modes will be briefly reviewed in the next section, with the purpose of gaining a better understanding of the social semiotic principles in exploring multimodal discourse.

### 6.2.3 Modality in modes other than language

As was mentioned above, modality resources are involved in all modes of communication. If we take a step further beyond the scope of language, the modality in other modes such as visual images, three-dimensional objects and sound can be examined through the lens of social semiotics. In the present section I shall provide a general delineation of the modality in those three modes of communication. Detailed elaborations of visual modality, which is the focus of the current research, will be presented in the ensuing section.

Modality in visual media has been theorized with the advance of social semiotics.
Hodge and Kress (1988: 128, 264) point out that modality judgements are conveyed through modality markers but interpreted through modality cues, that is, a more general category including both specialized modality markers and all the other bases for modality judgements in verbal as well as visual codes. The co-authors further summarize four general principles for analyzing visual modality: (i) The modality cues in visual texts are signifiers of semiotic activity, signifying the status of the text or parts of the text as high or low affinity; (ii) The modality value of a visual text is not fixed, but depends on receiver position and orientation, owing to the fact that receivers can be positioned in different ways in relation to mimetic content and to texts and producers; (iii) There is interaction between modality and ideology, ideology assigning modality values and modality legitimating ideological values; and (iv) Different genres establish sets of specific modality markers, and an overall modality value acts as the base-line for a specific genre with variations due to viewer/reader differences and other factors (Hodge and Kress 1988: 142).

Kress and van Leeuwen (1996, 2006) provide a more detailed analytical framework for analyzing visual modality. They have identified eight modality markers, which are means of visual expression involved in the judgement of visual modality, i.e. colour saturation, colour differentiation, colour modulation, contextualization, representation, depth, illumination, and brightness (Kress and van Leeuwen 2006: 160-163, see also van Leeuwen 2005: 167). Each of the eight aspects is regarded as a continuum that ranges from the maximal articulation to the minimum or absence of articulation. Explanations of these modality markers with instances from the data under examination will be given in the following section. It should be noted that modality value does not always increase with the amplification of articulation. As for the naturalistic standard, there is a certain point on each of the continuums that represents the highest modality value according to standard naturalism, that is, the ‘commonsense’, ‘everyday’ representation based on the resolution of the standard 35mm photographic technology. Nevertheless, besides this ‘naturalistic coding orientation’ (Kress and van Leeuwen 2006: 165) there are other contexts in which different kinds of truth are favoured. For instance, scientific
visuals aim at representing the essence, and thus the diagrams and figures with low articulations in colour, perspective, light and shadow are regarded as images with high modality value. This belongs to the ‘abstract coding orientation’. The concept of coding orientation will be explicated later in Section 6.3.

Taking a further step from two-dimensional images to three-dimensional visual displays, social semioticians investigate the modality of toys, sculptures, buildings and everyday design objects. It is found that some of the aforementioned principles work well in three-dimensional representations. For example, the modality of the toys for young children tend to be abstract and sensory in terms of detail and colour, whereas those for older children become more naturalistic and bear resemblance to the adult world. Nevertheless, three-dimensional modality differs from two-dimensional modality in the sense that there is no need to represent depth or the play of light and shade due to the three-dimensionality (Kress and van Leeuwen 2006: 252-255).

Moving from visual perception to aural perception, van Leeuwen (1999: 170-187) shows us that the modality of sound can be approached along the same lines as visual modality. Eight parameters are used to study the articulation of sound from different aspects, including pitch range, durational variation, dynamic range, perspectival depth, degree of fluctuation, degree of friction, absorption range, and degree of directionality. These eight dimensions are also seen as scales that run from the maximal degree to the minimal degree. In addition, the coding orientation adopted in a given communicative context determines the modality value of a particular sound, i.e. the degree and kind of truth that is assigned to the sound. Abstract-sensory, naturalistic, and sensory are the three types of coding orientations that are identified in people’s judgements of sound modality. The most abstract form of sound is music. Musical abstraction, or musical rationalization, does not necessarily require that the pitch and duration range should be reduced, which makes it possible for music to be both abstract and sensory.
6.3  Visual modality

I now narrow down the discussion to visual modality, with relevant examples from my data to account for the modality markers, i.e. the various parameters whose configurations express modality of a given visual event, and coding orientations, which provide criteria for modality judgement in different contexts.

6.3.1  Modality markers in naturalistic standard

There are eight means of visual expressions involved in the judgement of visual modality, which are termed as modality markers according to Kress and van Leeuwen (2006: 160-162). As was reviewed in Section 3.4.3, these modality markers are treated as continuums or scales running from the maximum potential to the lowest degree of articulation, with various degrees in between. At this stage I concentrate the discussion on the modality markers in the images with naturalistic coding orientation, which is the everyday standard as what can be observed with the naked eye, or technically speaking, as what is defined by the standard 35mm photography.

6.3.1.1  Colouration

As was mentioned above, along each of the continuums there is one point representing the highest modality. In terms of the photographic naturalism the point of highest value is not situated on either extreme of the continuum but on a certain point in between. Take the use of colour for instance. There are three dimensions of colouration to be considered, i.e. colour saturation, differentiation, and modulation. The continuum of colour saturation ranges from the use of maximally saturated colour to the absence of saturation, i.e. black and white. The naturalistic modality scale for colour saturation is represented in Figure 6-1 below.

![Figure 6-1 Modality scale for colour saturation (Kress and van Leeuwen 2006: 160)
Example (1) is taken from *Unit 1 Women of Achievement* in the senior secondary EFL textbook *New Senior English for China Student’s Book 4* (2004: 1), which provides concrete exempla for the explanation of modality markers. Four of the six small images in Example (1) are colour photographs (Images 2, 3, 4, and 6), and two of them are portraits (Images 1 and 5). In terms of colour saturation, Images 1 has the lowest modality due to its absence of colour saturation, whereas all the other five images possess the highest modality value in that they have the colour rendition which is either based on the 35mm colour photography (Images 2, 3, 4, and 6) or through the portraiture that aims at representing naturalistic colour rendition.

The second dimension is colour differentiation, which refers to the scale running from a maximally diversified range of colours to monochrome. As for Example (1), Images 3, 4, and 5 possess a greater diversity of colours than Images 2 and 6, while Image 1 has the lowest diversity among the six. The third aspect regarding colour that needs to be taken into account is colour modulation, ranging from the representation of all subtle nuances and full modulation of a given colour (e.g. the uses of different shades of green) to the use of plain, unmodulated colour. Take the skin colour of the represented participants in Example (1) for instance. Images 2, 3, and 6 represent the fine nuances as well as the effect of shades on the face complexion and arm skin, whereas Images 1, 4, and 5 apply relatively flat and less modulated colour for the depiction of skin.

### 6.3.1.2 Articulation of detail

When it comes to the articulation of details in an image, both background and foreground should be taken into consideration. Kress and van Leeuwen (2006) use the term ‘contextualization’ to describe the scale running from the most fully articulated and detailed background to the absence of background. In Example (1), the represented participants in Images 1, 5, and 6 are shown against the plain background of a single colour, which is almost equal to the zero articulation of background without any detail. The absence of setting, according to Kress and van
Leeuwen (2006: 161), lowers modality in that this kind of decontextualization represent the people within the picture frame as ‘generic’, ‘typical’ rather than specific figure who is associated with a particular point in time and space.

Images 2 and 3 are photographs with out-of-focus backgrounds, which is the intermediate degree between full contextualization and decontextualization. This meets with the naturalistic standard in which the combined effects of the 35mm photographic emulsion and the depth of field of standard lenses normally result in the less articulated background in comparison with what is represented in the foreground. If the background is represented as sharper and more defined, the impression of ‘artificial’ and ‘more than real’ will be produced (Kress and van Leeuwen 2006: 161). The background in Image 4 is considered as situated somewhere between the naturalistic contextualization and absence of background, in that the represented
participants (i.e. Jody Williams and a dog) are represented as being involved in an activity at a specific moment and in a particular location, while its pure black background, on the other hand, makes the whole image generic and thus seem unreal. Although this photograph might be taken on pitch-dark night, it is noteworthy that depiction of the background in this image causes it to deviate from the naturalistic standard.

As for the depiction of detail in the foreground, the term ‘representation’ is used to describe the pictorial details of represented participants (Kress and van Leeuwen 2006: 161). The scale of representation runs from maximum depiction of detail to maximum abstraction. It is evident that all the six images in Example (1) are of similar modality value in terms of detail articulation in the foreground. In these six images, the represented participants’ hair and facial features, as well as the creases in their clothes are depicted with the amount of details that meets with the naturalistic standard.

### 6.3.1.3 Depth, illumination, and brightness

‘Depth’ is concerned with perspective, and the scale of depth articulation ranges from maximally deep perspective to the absence of depth. Along this scale the highest modality value rests on the point that represents central perspective. This is commonly found in naturalistic photographs, in which the objects that are far from the lens tend to reduce in size and the real-life parallel lines converge at a vanishing point within or outside the picture frame. Therefore, central perspective is also termed as ‘one-point perspective’. The six images of Example (1) all belong to this category and meet the naturalistic standard. Kress and van Leeuwen (2006: 162) demonstrate some other points with different degrees of depth articulation such as angular-isometric perspective and frontal-isometric perspective. The modality values of these perspectives are lower than that of central perspective. They are named ‘isometric’ due to the fact that the measurements of a object, no matter how far it is from the lens, is represented with a three-dimensional view while the value of object scaling is still retained, and thus parallel lines remain parallel and there is no
vanishing point. The depth value of isometric perspective is then followed by that of simple overlapping. Towards the other extreme of the scale, the ‘more than real’ perspective, such as ‘fish eye’ perspective that swells in the centre of a sphere with converging lines at its sides, is also of low modality.

Another modality marker that can be scaled is ‘illumination’, which refers to the play of light and shade. The various degrees on this continuum range from the fullest representation of light and shade to its absence. As Kress and van Leeuwen (2006: 162) state, naturalistic depictions normally choose to represent participants as they are affected by a particular source of illumination. Although the light intensity differ among the six images in Example (1), they all employ naturalistic method in lighting depiction, thus having a high modality value in terms of illumination.

Last but not least, the modality marker ‘brightness’ is a scale that runs from a maximum number of degrees of brightness, to only the dark and light versions of a given colour. Kress and van Leeuwen (2006: 162) point out that the difference between the darkest and the lightest areas in an image can be as great as the contrast between deep black and bright white. Towards the other end of the scale, the difference can be minimal and thus a misty, hazy effect is produced. Likewise, what exceeds the naturalistic standard will be regarded as ‘more than real’, and hence lower modality value is assigned. Images 4 and 6 in Example (1) adopt a sharp contrast between pitch dark (i.e. the background and Jody Williams’ clothes in Image 4, and Lin Qiaozhi’s clothes in Image 6) and snow white (i.e. the dog next to Jody Williams in Image 4, and the background in Image 6). Nevertheless, this variation of brightness is still within the scope of photographic naturalism. Images 2, 3, and 5 are of higher modality value in brightness, due to the fact that they closely follow the naturalistic standard. On the contrary, the difference in terms of the degrees of brightness in Image 1 is quite small, and hence the vague, misty impression is produced and its brightness value is low.

It should be noted that within a given image, some modality markers may be amplified while others tend to be reduced. For instance, Image 1 in Example (1) is of low modality in terms of colour saturation, colour differentiation, colour
modulation, contextualization, and brightness, while the naturalistic treatments in terms of representation, depth, and illumination are employed in this image. The overall assessment of modality is generated by the viewer on the basis of the diversity of these modality markers (Kress and van Leeuwen 2006: 163). The effect of the different degrees of amplification or reduction in modality markers may result in a number of ‘modality configurations’, which cue a viewer’s judgements of modality with respect to ‘how real’ a certain image is to be taken (van Leeuwen 2005: 167). In the following section, I shall move on to the discussion of the modality values of different configurations in relation to the preferred visual truth in various contexts.

6.3.2 Coding orientation in other contexts

So far the focus has been on the naturalistic criterion, which is the most common and dominant standard on visual truth in our everyday life. However, those configurations with low modality in naturalistic standard can be of high modality value in other contexts. For example, colour\textsuperscript{1}, detail of texture, or play of light and shade is seldom found in the figures, diagrams and other technical drawings in scientific and technological discourse, and yet these images possess higher modality value than a naturalistic photograph as far as the scientific-technological context goes. Another exemplum comes from the advertising and commercial context. The images in this context are designed to arouse sensations, and thus they are generally sensory, ‘more than real’ in many aspects so as to produce the illusion of the benefits that the advertised products claim to bring. It can be inferred from the two examples above that the standard of visual modality is culturally defined. As Kress and van Leeuwen contend,

\[\ldots\text{visual modality rests on culturally and historically determined standards of what is real and what is not, and not on the objective correspondence of the visual image to a reality defined in some ways independently of it.}\] (Kress and van Leeuwen 2006: 163)
Drawing on Bernstein’s (1981) concept of ‘coding orientation’ (i.e. the sets of principles which inform the way in which texts are coded by specific social groups or within specific institutional contexts), Kress and van Leeuwen (2006: 165-166) identify three other reality principles along with the naturalistic one: technological coding orientation, sensory coding orientation, and abstract coding orientation. These are criteria for ‘what counts as real’ in different contexts or within different social groups. The naturalistic coding orientation has been explained in detail in the previous section, and now I shall turn to account for the three other categories with instances from the EFL textbooks under discussion.

6.3.2.1  **Technological coding orientation**

As for technological coding orientation, the reality principle lies in the ‘effectiveness’ or ‘usefulness’ of a given image as a ‘blue print’. The majority of the images with scientific or instructional purposes, including maps, architectural layouts, and instructions for dress-making, accord with this criterion (van Leeuwen 2005: 168). Example (2) is a map from my data showing the contours of the four countries of the United Kingdom, which is taken from *Unit 2 The United Kingdom* in *New Senior English for China Student’s Book 5* (2004: 9).

![Map of the United Kingdom](image)
In Example (2), most of the modality markers are reduced in articulation. Depth, for instance, is decreased to almost zero degree in the map. Besides, the use of colours does not represent the real colour of the land. Instead, they are merely conventional ‘colour codes’ for denoting different countries in the map so as to facilitate the viewer’s understanding. In the aspect of representation, it shows only the contours of the four countries, leaving out all the irrelevant details such as textures. However, these reduced articulations and flatness do not indicate low modality value. As is pointed out by Kress and van Leeuwen (2006: 164), in scientific-technological context what counts as real means what can be known by means of counting, weighing, measuring and other methods of science. Therefore, in the educational context like that of Example (2) where knowledge of geography is supposed to be instructed, the otherwise distractive details and articulations are reduced to low degrees.

6.3.2.2 Sensory coding orientation

Sensory coding orientation is primarily used in the contexts where pleasure principle dominates, such as advertizing, art, fashion, and interior decoration (Kress and van Leeuwen 2006: 165). It is often realized through high degrees of articulations that exceed the naturalistic standard, for example, high colour saturation, sharp detail in representation, high degree of illumination and brightness, etc. The background, on the contrary, is frequently represented as out of focus or simple sketches, so as to highlight what is represented in the foreground. Example (3) is taken from Unit 3 Let’s Paint in PEP Primary English Students’ Book I for Year 3 (2003: 27).
The image depicts a game that is designed for the purpose of practising the expressions of ‘Hi. How are you?’ ‘Fine, thank you.’ in primary EFL classroom. As is shown in the image, the characters (i.e. students in an English class) line up, one student greeting the next immediately after responding to the one before him/her. The represented participants are portrayed as anatomically distorted with unproportionately big head, thin limbs, large eyes, and wide smile. It is noteworthy that three out of the four represented participants in the image are depicted as ‘two-headed’ people, which aims at representing the dual role of one single student in taking turns to initiate as well as respond to greetings. Although this exaggerated representation may strike the viewer who holds naturalistic standards as ‘more than real’, the primary goal of sensory coding orientation is to create pleasure and arouse emotions. The background of the image is totally absent, and thus the represented participants in the foreground are even more outstanding and eye-catching. In short, all the pictorial techniques in the image are devised to be the sources ‘of pleasure and affective meanings’, and consequently Example (3) is regarded as an image of high modality in light of sensory coding orientation.

6.3.2.3 Abstract coding orientation

In abstract coding orientation, the standard for visual truth is the ‘essence’ or ‘general
truth’ that underlies the concrete and specific variations on the surface. This coding orientation is commonly adopted by sociocultural elites in the contexts of ‘high’ art and academia, where the reduction of diversified, individual differences to the general, essential quality is considered to be of high modality. In these contexts, the ability to identify and produce the texts that are based on this coding orientation is a mark of being an ‘educated person’ or a ‘serious artist’ (Kress and van Leeuwen 2006: 165). A case in point can be found in the EFL textbooks under examination. Example (4) is an excerpt from Unit 4 How do you get to school? in the EFL textbook Go for it Students’ Book I for Year 8 (2005: 22), which is a listening comprehension exercise for Year 8 students in junior secondary school.

(4)

Students are required to listen to a dialogue and choose the transportation means by which Nina gets to school. One of the premises for doing this exercise is the correct reading of the abstract visual symbols that stand for locations including Nina’s home (i.e. ), school (i.e. ), as well as transportation means such as by bike (i.e. ), by subway (i.e. ), by bus (i.e. ), and on foot (i.e. ). In these visual symbols, most of the details that represent individual nuances are reduced to the abstract truth. For example, the human figure is identifiable as a person who is walking, but a viewer cannot tell the gender or age due to the lack of colour, brightness, illumination, and representation of details and depth. Other symbols that represent destinations and transportation means also employ decreased articulation,
for instance, low colour differentiation, and low brightness variation with misty impression, etc. In sum, as is pointed out by van Leeuwen (2005: 168), from the point of view of the essential or general truth, all details that indicate specific, individual features are irrelevant. The more an image represents the general pattern that underlies the superficially different instances, the higher its modality will be according to abstract coding orientation.

6.4 A multilevel view on visual modality in EFL textbook discourse

Some consideration has so far been given to the eight visual parameters that serve as modality markers in judging what counts as real in a given context or among the members of a certain social group. The reality principles that have been covered include the everyday naturalistic one as well as the technological, sensory, and abstract coding orientations adopted in different contexts. I now in a better position to embark on a comparative study of coding orientation in texts for different pedagogic context, with the purpose of providing a multilevel view on the visual modality configurations in primary and secondary EFL textbook discourse. The data that will be drawn upon are three teaching units concerning animals. These three units are comparable in the sense that they share the same field, i.e. the ‘topic or focus of the activity’ in which communicative parties are engaged, or more broadly speaking, the ‘institutional focus or social activity type’ (Eggins 2004: 9, 103; see also Martin 1984: 23, 1992: 536). In addition, all the EFL textbooks under attention are similar in the register variable of mode, in that they are all edited to be read as formal instructional materials and they all draw upon both linguistic and visual meaning-making resources.

6.4.1 Sensory coding orientation in We Love Animals

The unit concerning animals in primary EFL textbooks is Unit 4 We Love Animals in PEP Primary English Students ‘Book I for Year 3 (2003: 38-47). As was indicated in Chapter 5, cartoon is the dominant image style in primary EFL textbooks. I shall
take the cartoons in the teaching section *Culture* (see Example (5) below) as an example to examine the coding orientation in the EFL textbooks for primary education.

(5)

![Example (5)](image-url)

[PEP Primary English Students' Book I for Year 3 2003: 47]

There are four images in Example (5), each of which depicts one animal that is considered to be typical in a given country (i.e. *panda* for *CHINA*, *beaver* for *CANADA*, *eagle* for *USA*, and *kangaroo* for *AUSTRALIA*), with the name and contour of the country as its background. The visual realization in each modality marker varies in the degree of articulation. Specifically, in terms of colouration, high colour saturation as well as medium colour differentiation and modulation are employed. As for representation, the foregrounded animals are portrayed as ‘more than real’, in the sense that their emotions and temperament are added into the personified visual displays via facial expressions and gestures. In real life normally it is usually not easy to tell how an animal feels without certain expertise, but the personification adopted here overtly reveals to us something of the animals’ emotions. To be specific, the panda is represented as outgoing with waving gestures and a wide smile; the beaver seems happy but somewhat cocky with arms akimbo; the red-eyed eagle appears a bit fierce with wings spread; and the kangaroo looks funny but caring, with a superman’s cloak on its shoulder and a baby kangaroo in its pouch. In addition, the
animals are represented as disproportionately big as compared with the country contours in the monochrome background, which highlights and further enhances the entertaining effect of the represented participants in the foreground. As pictorial techniques used in these images follow the pleasure principle whose purpose is mainly for fun, the criterion of truth employed in Example (5) can be recognized as sensory coding orientation.

As was analyzed above, the ‘exaggerated’ and ‘sensational’ style is actually considered to be the criterion of truth for primary school students. Textbook editors are by no means school-aged children, and normally they hold naturalistic coding orientation as criterion of truth. Obviously they have chosen to adjust their truth criterion to the sensory one in designing the visual display for primary EFL textbooks. As Hodge and Kress (1988: 151-153) point out, when there is instability between the ‘world of the producer (WP)’ and ‘world of the reader (WR)’, either the efforts must be made to resolve the difference between WP and WR or that difference will be the basis for antagonism. The efforts may come from producer or receiver or both, and in the case of the primary EFL textbooks under discussion, the effort to resolve the difference is made by the WP, i.e. textbook editors. The sensory visual style in primary EFL textbooks is presumed to attract the attention and excite the imagination of the school-aged English learners. To sum up, sensory coding orientation in the visual images of primary EFL textbooks creates a virtual world in which the truth criterion is pleasure principle. This visual arrangement is supposed to be of high modality value to primary school students, as it is discerned by the textbook editors who endeavour to establish a strong solidarity relationship with the school-aged textbook readers.

6.4.2 Abstract-sensory coding orientation in Why Do You Like Koalas

There is a strikingly similar instance in the junior secondary EFL textbooks under attention, which is an exercise section entitled Match the animals with the countries (see Example (6) below) in Unit 3 Why Do You Like Koalas of Go for it Students’ Book II for Year 7 (2005: 13-18). Both the primary school text and the junior
secondary school text include images depicting animals and the relevant contours of countries. Their difference lies in the fact in Example (5) the contours of countries serve as the background, whereas in Example (6) the images of animals and countries are divided into two sets. The lower set represents three animals (i.e. lion, panda, and koala), each of which is deemed as the famous, representative animal of a certain country. The upper set depicts three countries with names as captions, i.e. China, Australia, and South Africa. Here the junior secondary students are no longer provided with the correct match between animals and countries. Instead, they are required to match the animals with the corresponding countries by themselves. Taking a multilevel view on the teaching units that are concerned with the same topic in primary and junior secondary EFL textbooks, analysts may find out that the repeated use of similar teaching materials complies with the requirement of ‘high reappearance rate for language teaching materials’, as is stipulated in the Curriculum Standards for English. Furthermore, the transition from simple to complex instructional designs reflects the principle of ‘proceeding in an orderly way and advancing step by step’ (Zhonghuarenmingonghegovu jiaoyubu 2001a: 48).

I now begin the examination of the coding orientation in Example (6) with the analysis of the three animal images. It should be noted that reduced articulation is employed in the aspects of representation, contextualization, illumination, and depth. To be specific, simple line drawing is adopted to capture the essence of the animals, that is, what makes them identifiable as lion, panda, and koala. For example, the long heavy mane around the top and sides of the neck, which is the symbol of a male lion, allows the viewer to easily recognize this animal. Nonetheless, the shapes of the three animals and the textures of their fur are brought down to the essential quality. Unlike Example (5), these animals are not depicted without background. Instead, they are represented as snapshotted in their actual living environment. For instance, the koala is described as clinging to a eucalyptus tree. Nonetheless, the background is low articulated and merely represents the general scenario of the environment, leaving out all the details such as the texture of tree bark or the weather condition which would otherwise be indicated in naturalistic photographs. The abstract style is
also reflected in the low articulation of illumination and depth, in that there is no play of light and shade or perspectival foreshortening in the images. The three maps above the animal images further enhance the sense of abstractness. In these maps most of the modality markers are greatly reduced in articulation. Although colour differentiation can be found, the purpose of using various colours in the maps is to distinguish the countries under discussion from the adjacent countries and oceans. Underlying the abstract modality of Example (6) is the truth criterion which holds that the more an image arrests the essential quality or the general truth of what is represented, the higher the modality value will be.

(6)

[Go for it Students’ Book II for Year 7 2005: 15]

However, it is noteworthy that there is a tension between the abstract modality and the emotive orientation in these images. Unlike the typical abstract images such as scientific diagrams, the animal images in Example (6) draw on the pleasure principle to some degree. A high degree of colour saturation is applied, and subtle nuances of a given colour are identified. Take the green colour for instance. The variations of greenness include the bright green for twigs and new leaves, the brownish green for meadow, and dark green for leaves on the eucalyptus tree. The use of colour here is not merely for telling apart different components, nor does it aim at faithfully reproducing the real colours in natural environment. The colouration
here is ‘a source of pleasure and affective meanings’ (Kress and van Leeuwen 2006: 165), whose purpose is to make the images pleasant to the eyes and to produce an affective effect that helps attract attention and maintain interest.

It can be inferred from the analysis above that the criterion for truth in the junior secondary EFL textbooks is both abstract and sensory. In other words, the coding orientation of Example (6) is abstract-sensory. I attempt to explain the adoption of this abstract-sensory coding orientation in the junior secondary educational context by looking at the special role of abstract modality in education and the pleasure principle that continues into adolescent education. Abstract modality goes beyond the surface to capture the underlying essence (c.f. realism in art and literature as discussed by van Leeuwen 2005: 168). It is used among sociocultural elites (Kress and van Leeuwen 2006: 165), and understandings about producing and reading images with an abstract coding orientation is part of the education of ‘cultivated’ people. An abstract coding orientation is more evident in textbooks on natural science where diagrams and figures are widely applied. As was shown above, the tendency towards the ‘conceptual’, ‘abstract’ coding orientation gains momentum in junior secondary context, while the sensory style is still retained. The tension between the educational purpose and pleasure principle accounts for the coding orientation in the junior secondary EFL textbooks, which is ‘abstract’, ‘conceptual’ with respect to representation and contextualization, but pleasure-eliciting in terms of colouration.

6.4.3 Naturalistic coding orientation in *Wildlife Protection*

The teaching unit concerning animals in junior secondary EFL textbooks deals with a more serious topic, i.e. *Wildlife Protection*, which is the fourth unit in *New Senior English for China Student’s Book 2* (2004: 25-32) for Year 10 students. I concentrate the discussion on the *Warming Up* section (see Example (7) below) of the unit, which describes three animals (i.e. *panda*, *milu deer*, and *South China tiger*) that were once endangered in China due to lack of food supply or over-hunting. The corresponding verbal texts also introduce some measures that have been taken to protect these species from extinction and the resulting achievements.
Images of the three animals are inserted into a table that shows the causes of the near extinction, the animals’ habitats, as well as the contrast between the previous worsening condition and the current improved situation. These images are photographs with naturalistic coding orientation, representing the real-life situation of the wild animals in natural environment, i.e. two pandas eating in a bamboo grove, a number of milu deer drinking by a river, and a South China tiger running on meadow. The relevant truth criterion here is the naturalistic standard.

As noted above, the naturalistic coding orientation is the shared, dominant truth criterion for the general public. Images with naturalistic coding orientation represent what is snapped into the photoframe the way we see it with the naked eye in reality. When it comes to public concerns such as the pressing environmental problems and corresponding solutions as is demonstrated in Example (7), the adoption of cartoons might downplay the seriousness of the issue, while the application of schematic
drawings that represent animals as symbols or specimens may result in distancing the viewer from the endangered species that are urgently in need of help. Naturalistic photographs, on the other hand, represent the endangered wildlife animals in the way that is in accordance with the cultural norm. They faithfully reproduce the living conditions of these endangered animals, hence creating the sense of being personally on the scene. In other words, these photographs follow the coding orientation that all members of a given culture share, regardless of their educational levels or special training (Kress and van Leeuwen 2006: 165-166). The naturalistic coding orientation here positions the viewer as involved in the environmental protection campaign in which all members in the shared culture are supposed to participate.

So far I have considered the ‘interdependence’ (Hodge and Kress 1988: 161) between modality and social relations from a multilevel perspective by examining the multimodal texts about animals in the context of EFL education. The above analysis and discussion illustrate that modality is ‘pervasive, appearing everywhere in an utterance or text’ (Hodge and Kress 1988: 127), and the resources that convey modality value in a text are multimodal. In the ensuing section, I shall focus on one type of visual style, i.e. cartoon, and discuss the appropriateness of its proportion in the educational setting.

6.5 Cartooning and childist ideology

As was demonstrated in the above analysis, the criterion for truth adopted in the primary and junior secondary textbooks stand in stark contrast to those adopted in the senior secondary textbooks, in that the element of sensation is highlighted to a greater extent in the former. Considering the high proportion of cartoon employed in primary and junior secondary textbooks (97.9% and 89.6% respectively; see Table 5-1 in Section 5.3 of Chapter 5 for the statistics), it is evident that the sensory coding orientation has a pervasive influence on the EFL textbooks for children and teens. In the present section, I seek to find out what kind of educational ideology lies behind this high proportion of cartooning and discuss whether it is justifiable or excessive.
Before a close reading of our data, it is necessary to distinguish the cartoons for children (e.g. those used in primary and junior secondary EFL textbooks) from those targeting adult viewers (e.g. Examples (8) and (9) below). It is suggested in the present discussion that one of their major differences lies in their functions, or the different purposes of use in specific contexts. Examples (8) and (9) are grouped as ‘satire cartoons’, among many other types of cartoons for adults, which are commonly found in newspapers and magazines. It can be inferred that one of their social functions is to convey particular ironic messages to the target readers or consumers, who are chiefly and normally adults. In this case, both image producer and viewer hold naturalistic coding orientation as their criteria for truth in everyday life. In contrast, those cartoons used in children’s picture books are without such implications of irony. Rather, cartoons for children represent the world as what is assumed to be real for children. In other words, the image producers choose to surrender their own normal criteria for truth to cater for what children might accept as real. In short, although both types of cartoons may draw on sensory coding orientation, the reasons why they choose to represent the world as unreal differ: those for adults are to evoke sensations and convey implications, whereas those for children are yielding to another criterion for truth that is supposed to be possessed by the viewer group of children.

(8)

As was analyzed in Section 5.3.1.1 of Chapter 5, happy and satisfied faces manifest one of the major attitudinal orientations in cartoon, which is the dominant visual style in the primary and junior EFL textbooks. The inscribed happiness, as was explained in Section 5.3.1.1, indicates a positive attitude toward English learning. The overwhelming cartoon style, as will be explicated in this section, uncovers the underlying ideology of ‘kids must play to learn’ childist ideology. As Martin (1985b: 54-57) points out, this ideology holds that children are ‘individuals’, ‘spontaneous learners’, ‘cognitively immature’, ‘creative’, ‘innocent’, ‘egocentric’, ‘imaginative’, ‘ignorant’, and ‘irresponsible’. In pedagogic setting, this attitude towards children, as Martin (1985b: 57) comments, is by no means conducive to the empowerment of children. I will first clarify the reason why it is the excessive cartooning rather than the inscribed emotion that embodies this attitude towards children, which will precede a discussion on some themes in the twentieth-century educational theories that may reinforce childist ideology, in relation to the sociological and systemic functional semiotic position that the current study assumes.
6.5.1 Embodiment of childist ideology

Admittedly in the majority of images in the EFL textbooks under discussion, cartoons and inscribed happiness often come together while photographs usually invoke seriousness. It seems justifiable to contend that the conjunction of cartooning and happiness reflects one dimension of ‘kids must play to learn’ childist ideology. In order to explore this further, let’s examine some exceptional instances where cartooning and happiness do not go together.

The first example is a cartoon without inscribed happiness (see Example (10) below), which has been drawn on in Section 4.3.1 of Chapter 4 to analyze the ENGAGEMENT device of labelling. In this cartoon, the inscribed feelings are worry rather than happiness, arising from the story that the represented participants (i.e. Amy and her mother) cannot find the white socks.

Without the inscribed happy emotion, a viewer will still find this image not that serious due to the fact that cartoon style indicates that the image was produced for fun. It can be inferred that the underlying assumption here is ‘we need to present the world as playfully as possible, otherwise our children readers will not find it interesting to learn the vocabulary’. This reflects the relatively pervasive view which assumes that children ‘have trouble distinguishing between fantasy and reality’ and would rather ‘live in a world of dreams or play’ (Martin 1985b: 55). In other words, the editor chooses to adjust their truth criterion to the children’s so as to please the primary school students. To put it in Hodge and Kress’ (1988: 151-153) term, ‘world of the producer (WP)’ yields to ‘world of the reader (WR)’ in addressing the inconsistency between adult truth criterion and what is assumed to be fascinating to the children.
The second group of instances (see Examples (11) and (12) below) differs from Example (10) in terms of both visual style and Attitude, in that they are naturalistic photographs with positive affectual meanings inscribed. Example (11) is taken from Unit 6 I’m more outgoing than my sister in the junior secondary EFL textbook Go for it Students’ Book I for Year 8 (2005: 35). The multimodal text snapshots three teens (i.e. James Green, Huang Lei, and Mary Smith) talking about whether or not they think friends should have a lot in common. Although the three teens have different views on the issue as are indicated in the verbal texts (e.g. ‘I like to have friends who are like me…’ ‘It’s not necessary to be the same. I like to have friends who are different from me…’ ‘I don’t really care…I don’t think differences are important in a friendship.’), what the three photographs have in common are the smiling faces that indicate happy, satisfied, and confident affectual meanings. Example (12) is an anecdote excerpted from Unit 5 First aid in the secondary EFL textbook New Senior English for China Student’s Book 5 (2004: 38). It reveals that the seventeen-year-old John Janson saved his neighbour’s life with first aid knowledge and skills, and thus was conferred the Lifesaver Award. The anecdote was accompanied by a photograph capturing John Janson’s grinning face. In Examples (11) and (12), the represented participants’ positive Attitude is oriented towards life in general (i.e. making friends and helping others). Normally a viewer will not group them as childish images despite the depiction of happy teenagers. That is because these images adopt the
Contestable Reality: A Multilevel View on Modality in Visual Communication

everyday, dominant standard of visual truth (i.e. naturalistic coding orientation) in everyday life.

(11)

(12)

The comparative analysis above shows that when the inscribed happiness is
taken away, childist ideology is still present. Therefore, it can be inferred that the ‘kids can only learn by playing’ childist ideology primarily manifests itself in the playfulness of cartooning not through the inscribed happiness in the depicted children’s faces. To sum up, two points need to be clarified: first, the multimodal resources that encode ATTITUDE in EFL textbook discourse, as analyzed in Chapter 5, realize the meanings of positive emotions, high moral standards, and a healthy aesthetic interest that students are supposed to develop in their study and life; second, I am not suggesting that cartoons should all be banned in EFL textbooks. Rather, it is an excess of cartooning that should be cautioned against, because it runs the risk of underestimating students’ ability to learn language from texts that deal seriously with a range of contexts. The latter point will be further explored the ensuing section, with reference to a review and reconsideration of educational movements in the twentieth century.

6.5.2 Relevant educational theorizing

When it comes to the question of where the aforementioned childism comes from, namely, what kind of educational theory behind the assumption that children can only play to learn, it is necessary to trace back to some trends in the twentieth-century educational theory such as ‘constructivism’, ‘progressivism’, and ‘child-centredness’.

Constructivism as an educational movement is strongly associated with Piaget’s psychological conception on stages of cognitive development. Two of Piaget’s (1976, 1977) important arguments are relevant to the current discussion. First and foremost, he stresses that the child’s cognitive structures are distinct from those of the adult, and believes that children develop their cognitive structures purely via their independent explorations of the material environment. Secondly, he concentrates on the gradual, universal stages through which children develop cognitively at their own pace, and argues that children are unable to successfully conduct specific tasks until they are cognitively prepared at appropriate developmental stage. For example, the first stage among the four universal series, according to Piaget (1977: 277), is the ‘sensorimotor’ period that is characterized with ‘egocentric perceptions’. The child
or infant at this stage is alleged to be incapable of making ‘any distinction between his internal state and things outside’ (Piaget 1977: 207). It is claimed that the child’s cognitive development of logical thinking involves the ‘decentring’ or the distinction made between self and non-self, which cannot be fully achieved until adolescence (Piaget 1976). However, research on neonatal behaviours shows that an infant is capable of not only ‘differentiating himself from the world’, but also ‘distinguishing things from persons’ (Trevarthen 1980: 324; see also Field and Fox 1985; Messer 1994). Furthermore, it has been found that infants interact with and are responsive to caregivers and people around him (Murray and Trevarthen 1985), and thus they ‘possess an intrinsic and growing motivation to gain knowledge from others’ (Trevarthen 1980: 336).

The Piagetian orientation emphasizes the role of a materially rich environment in which children are left to make their own discoveries. This position has been very influential in educational theory and practice, especially in the field of science education. Although there are a variety of interpretations concerning the constructivist mode, ranging from ‘radical constructivism’, ‘didactic constructivism’ to ‘pragmatic social constructivism’ (Larochelle et al. 1998: viii), they basically imply that ‘knowledge is always knowledge that a person constructs’, and stresses ‘the need to encourage greater participation by students in their appropriation of scholarly knowledge’ (Larochelle and Bednarz 1998: 3). Nevertheless, as is true of its applications in education, the theory of constructivism has been criticized. One of the important and relevant critiques is from within psychology on the validity of the experimental data from which the interpretation of the ordered stages derives. It is contended that children’s cognitive abilities were constantly underestimated in Piaget’s experiments, in that he did not design appropriate tasks that children could carry out, but simply blamed the children’s immature cognitive development for their failures in completing the tasks (Borke 1975; Donaldson 1978; Samuel and Bryant 1984). Another pertinent critique is concerned with the neglect of social and semiotic dimensions in the theory. As is suggested by Sinclair (1974: 58), for Piaget ‘the subject himself is the mainspring of his development, in that it is his own activity
on the environment or his own active reactions to environmental action that can make progress.'

This persistent emphasis on the individual child’s engagement with the material environment as the determining factor and the disregard of the child’s interaction with other communicating persons, as is pointed out by Painter (1999: 15-16), downplay the role of language in children’s cognitive development and reflects ‘a lack of awareness of the relation between text and context’ as well. Moreover, as Vygotsky's (1978: 88) social approach to cognitive development indicates, ‘human learning presupposes a specific social nature and a process by which children grow into the intellectual life of those around them’. Within educational context, a distinction should be made between what a child can possibly achieve with adult’s aid (i.e. the ‘actual developmental level’) and what he/she can achieve without assistance from others (i.e. the ‘level of potential development’). The distance between these two levels is termed as the child’s ‘zone of proximal development’ (Vygotsky 1978: 86), in which effective instruction can be conducted. Accordingly, the child’s cognitive development advances not in a unilinear sequence but in a ‘spiral’ way from no mediation to external mediation with signs, then to the internal mediation at a higher level (Vygotsky 1978: 52-57; see Painter 1999: 25-30 for further discussion).

Still another critical and relevant comment is made on the ‘de-emphasis’ of the role of teacher (and educator in general including textbook editor) in constructivist theory, because it suggests that children cannot learn until they are cognitively ready to learn. For example, ‘radical constructivism’ (von Glasersfeld 1998) argues that ‘knowledge cannot be transmitted, but must instead be constructed by each student individually’. Therefore, for von Glasersfeld (1998: 27), what a teacher can do is simply ‘to “orient” students’ efforts at construction’. Another influential movement that seems to accord with the diminishment of educator’s role arises from the progressive, child-centred approach. Although it is argued that in primary classroom, the ‘open’, ‘child-centred’ pedagogy might no more free the child from external controls than traditional pedagogy (Alexander 2000: 562; see also Sharp and Green 1975), its impact on the teaching of subject English, as Christie (2004: 177) observes,
results in the teacher’s diminished role as merely ‘trusted adult’ who facilitates or mediates the child’s personal growth (e.g. Dixon 1967). In the constructivist, progressive model, the nature of knowledge becomes relativistic and the authority of teacher in the pedagogic relationship of schooling is confused (Christie 2004: 173; see also Muller 2000).

Returning to the discussion of the visual images in our data, one can identify the practice of foregrounding the students while downplaying the significance of educator at the expense of the value of knowledge. As the For Teachers in PEP Primary English Teachers’ Books state, one of the major characteristics of the series of primary students’ EFL textbooks is ‘highlighting interest arousal’, which is further specified as ‘maximally arousing students’ motivation for and interest in learning, which is reflected in content, form, methodology, illustration and design of the textbooks… The whole series of textbooks adopts cartoon illustrations with bold colouration, integrating verbiage and image and presenting exquisite design.’ (2003: II) However, the pervasive childish cartooning positions the primary and junior secondary students in an allegedly naive ‘world of their own’, and deliberately distances them from the real, logical world in which students need to establish their social identity. Three points need to be emphasized to guard against possible misreadings of the position adopted here. Firstly, it is by no means argued that books for children should not be illustrated with cartoons, nor do I intend to undermine the role of sensory visual style in attracting attention and exciting imagination. In certain contexts, a fairy tale or fictional story for example, it may be appropriate to illustrate the text with exaggerative, sensational visual images. Furthermore, as was noted above, sensory coding orientation can also be applied in advertising and satire cartoons that target an adult audience. What matters here is the issue of context. It is the educational context in which textbook discourse is embedded that prompts us to reconsider whether the visual images in such context should be purely for fun.

Secondly, what is asserted here is not the prohibition of cartoons in textbook discourse. Rather, it is the proportion of visual styles that should claim scholarly attention. The overdose of fanciful cartoons (97.9% in primary textbooks and 89.6%
in junior secondary ones) is a manifestation of ‘invisible’ (see below for further discussion) pedagogy (Bernstein 1975). As Christie (2004: 178) points out, the pedagogic relationship should be viewed as ‘both constraining and enabling’. Curriculum theorizing influenced by constructivist and progressivist model of knowledge and the learner may ‘diminish the status of knowledge structures, as well as the role of the teacher’ (Christie 2007: 8, see also Christie 2002, 2004; Muller 2000).

Thirdly, the identification of the constructivist, progressivist influence that is reflected in the textbooks is not equal to the position that views constructivism or progressivism as the theoretical foundations of Curriculum Standards for English. As Cheng and Gong (2005) point out, the current Curriculum Standards for English on which the textbooks are based belongs to ‘multi-syllabus’ or ‘mixed syllabus’ that integrates a variety of characteristics of different syllabi. Moreover, they contend that in the curriculum standards knowledge and skills are valued not only as teaching content but as teaching goal (see Section 2.3.2 for the discussion of the ‘three-dimensional goal’ in the current curriculum). In other words, what the current discussion focuses on is whether or not the multimodal resources in textbooks live up to what is asserted by educationists, rather than intending to subvert the curriculum standards.

The remainder of this section will try to offer an alternative, sociological stance towards pedagogic discourse and nature of knowledge. Pedagogic activities and the structuring of knowledge have been the major concerns of the British sociologist Basil Bernstein (1924-2000). As Bernstein (1990: 165) points out, what was lacking in the sociology of education from the 1970s was ‘the analysis of the intrinsic features constituting and distinguishing the specialized form of communication realized by the pedagogic discourse of education’. The preoccupation on the power relations external to knowledge (i.e. class, gender, race, etc.) leads to negligence of the intrinsic relations within knowledge itself, i.e. its nature and structuring. Bernstein’s approach to pedagogic discourse and knowledge begins with his adaptation of the linguistic notion of code in addressing cultural transmission (e.g. Bernstein 1981),
which has later developed into his conceptualization of pedagogic device and knowledge structure (Bernstein 1990, 2000). Code in Bernsteinian theory can be defined as ‘a regulative principle, tacitly acquired, which selects and integrates relevant meanings, forms of realizations and evoking contexts’ (Bernstein 1981). One concern that is relevant to the current discussion concerning educational context is that, a decision needs to be made between the following two choices: either changing the underlying structuring principles of curriculum to cater for the code already possessed by the students, or providing the students with the code that is essential in enabling success (Maton and Muller 2007: 17). This opposition is closely related to the distinction between ‘invisible’ and ‘visible’ pedagogies (Bernstein 1975: 116-151). In an invisible pedagogic practice, the rules of regulative and discursive order are implicit and known only to the transmitter (e.g. teacher) but remain implicit or ‘invisible’ to acquirers (e.g. students), with emphasis placed on ‘acquisition-competence’; whereas in a visible pedagogic practice, the rules of regulative and discursive order are explicit and known to both transmitter and acquirers, with emphasis put on ‘transmission-performance’ (Bernstein 1990: 70-71). The basic difference between them lies in ‘the manner in which criteria are transmitted and in the degree of specificity of the criteria’(emphasis in the original) (Bernstein 1975: 116-118).

According to Bernstein (1990: 183-184), pedagogic discourse is defined as ‘a principle for appropriating other discourses and bringing them into a special relation with each other for the purposes of their selective transmission and acquisition’. It involves ‘regulative’ and ‘instructional’ discourses, with the former ‘embedding’ (Bernstein 1990: 183) or ‘projecting’ (Christie 2002, 2004; Martin 2007) the latter. Instructional discourse refers to ‘a discourse of competence (skills of various kinds)’ and their relation to each other; whereas regulative discourse means ‘a discourse of social order’, which creates ‘specialized order, relation and identity’. In the EFL educational setting under attention, the overall goal of ‘integrated language-using competence’ (see Section 2.4.3 for detailed discussion) can be considered as the instructional discourse, while the social identity valued in society is generated through
regulative discourse, which is of primary significance and should be regulated via educator as the agent of symbolic control. As Christie’s (2004: 179-180) observation on the authority of teacher’s role in classroom demonstrates, the regulation is performed ‘by selection and appropriation’ of instructional discourse. It can be inferred that educator including textbook editor, like the teacher in classroom as Christie comments, is the authority that mediates the processes by which regulative discourse organizes pedagogic activity and its evaluation, which in turn makes possible the competence related to instructional discourse.

For Bernstein, pedagogy works through the operation of a pedagogic device, which consists of three sets of rules that integrate power, knowledge, and consciousness: ‘distributive rules’ that distribute forms of consciousness through the distribution of forms of knowledge, ‘recontextualizing rules’ that regulate the construction of instructional discourse for specific pedagogic purpose, as well as ‘evaluative rules’ that establish the criteria for legitimating the forms of knowledge communicated. Each of these rules has to do with a specific field of activity and is associated with typical institutionalized sites (Christie 2007: 7; Maton and Muller 2007: 19). The three rules and their related fields together represent an ‘arena of struggle’ (1990: 206), which is summarized in Table 6-1 below.

<table>
<thead>
<tr>
<th>Field of practice</th>
<th>Production</th>
<th>Recontextualization</th>
<th>Reproduction</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Form of regulation</strong></td>
<td>distributive rules</td>
<td>recontextualizing rules</td>
<td>evaluative rules</td>
</tr>
<tr>
<td><strong>Kinds of symbolic Structure</strong></td>
<td>knowledge structure</td>
<td>Curriculum</td>
<td>pedagogy and Evaluation</td>
</tr>
<tr>
<td><strong>Principal types</strong></td>
<td>hierarchical and horizontal knowledge structures</td>
<td>collection and integrated curricular Codes</td>
<td>visible and invisible pedagogies</td>
</tr>
<tr>
<td><strong>Typical sites</strong></td>
<td>research papers, conferences, laboratories</td>
<td>curriculum policy, textbooks, learning aids</td>
<td>classrooms and examinations</td>
</tr>
</tbody>
</table>

Besides identifying the pedagogic device, which is the condition for the
construction of pedagogic discourse, Bernstein (2000) also explores the nature of pedagogic discourse or knowledge structure. Bernstein’s theorization of knowledge begins with the distinction between ‘horizontal’ and ‘vertical’ discourses. Horizontal discourse is the everyday or ‘common-sense’ knowledge, which ‘entails a set of strategies which are local, segmentally organized, context specific and dependent’ (1999: 159). Vertical discourse on the other hand either ‘takes the form of a coherent, explicit, and systematically principled structure’, which is termed as ‘hierarchical knowledge structures’ as in sciences, or ‘takes the form of a series of specialized languages with specialized modes of interrogation and specialized criteria for the production and circulation of texts’ (ibid), which is referred to as ‘horizontal knowledge structures’ as in social sciences and humanities⁴. To sum up, one of the major contributions of Bernstein’s approach, as Maton and Muller (2007: 29) articulate, ‘lies in its capacity to render knowledge visible as an object of study’.

Education has long been one of the major contexts under exploration by systemicists and social semiotics, and this has given rise to a fruitful transdisciplinary dialogue between Bernsteinian sociology and systemic functional semiotics. In the pursuit of the linguistic features of pedagogic discourse and nature of knowledge, the linguists and educationists who take the systemic functional approach have engaged in studies of genre (Martin and Rose 2008), technicality and abstraction (Martin 2007; Halliday and Martin 1993), exchange structure and appraisal (Martin 1992, 1999; Martin and White 2005), and multimodality (O’Halloran 2005; Unsworth 2001, 2008). All of those dimensions contribute to the understanding of the relationship between ‘disciplinarity, cumulative learning, and language and literacy’ (Freebody et al. 2008: 194).

6.6 A note on visual style within a macrogenre

So far I have examined the criteria for truth in different visual styles in EFL textbook discourse, and discussed the appropriateness of the amount of one type visual style (i.e. cartoon) in educational setting. The social definition of truth, as I have been
arguing to this point, is constructed through the affinity and antagonism of social
groups, and it reflects the shared value and alignment among members of a given
institutional or social group. Social relations are ‘constructed and mediated through
semiotic activity’ (Hodge and Kress 1988: 161). Nonetheless, it should be noted that
members of a social group that normally share the same coding orientation by no
means only have access to one style of visual display. With reference to the EFL
textbooks under discussion, I find that various visual styles are employed in different
elemental genres that constitute the macrogenre of a teaching unit. I shall attempt to
analyze and account for this phenomenon by examining different visual styles that are
adopted in two reading materials within the same teaching unit.

6.6.1 The teaching unit as macrogenre

The term ‘genre’ has been approached in a great variety of disciplines, while in the
present discussion I focus on what it means in linguistic sense and draw on the
approach to genre that is developed by the ‘Sydney School’5. Within SFL, genre is
interpreted as a staged, goal-oriented social process (Martin 1984). Genre is the
contextual semiosis above the semiotic stratum of register that reflects the ‘variations
in culture’ (Martin 1992: 504-507; Martin and Rose 2008: 16)6.

Genre analysis can be applied both to much longer texts that spread across
several pages and to ‘elemental’ genres realized as relatively short texts. Martin
(1994) proposes the concept of ‘macrogenre’ to deal with longer texts in which
various constituent genres can be identified. It is argued here that each teaching unit
in the EFL textbooks under discussion can be treated as a macrogenre, in that the
entire unit is used to accomplish a specific educational goal and is divided into a
number of staged sections. Each of the sections belongs to a specific genre. Take
Unit 2 The Olympic Games in New Senior English for China Student’s Book 2 (2004:
9-16) for instance. As is indicated in the corresponding New Senior English for
China Student’s Book 2 (2007: 29), the goal of that teaching unit encompasses four
categories, i.e. ‘topic (Ancient and modern Olympic Games; Olympic spirit),
vocabulary, function (Talking about interests and hobbies), and grammar (simple
future tense). To achieve this goal, the whole unit consists of eight teaching sections, i.e. Warming Up, Pre-reading, Reading, Comprehending, Learning about Language, Using Language, Summing up and Learning Tip, and these sections represent different types of genres. For instance, the Reading section (see Example (13) in Section 6.5.2 below) in that unit is an description of modern Olympic Games that takes the form of an interview; the reading material of Using Language (see Example (14) in Section 6.5.3) is a narrative concerning two figures in Greek mythology; while Learning Tip (see below) of that teaching unit is an exposition that explains why the students should be active in pair work or group work.

```
[New Senior English for China Student’s Book 2 2004: 16]
```

Martin and his colleagues (e.g. Martin 1994; Martin and Rose 2008: 216-223) draw on Halliday’s concept of logico-semantic relations to account for the relations between componential genres within a macrogenre. The examination of the relations between different sections within a teaching unit is beyond the scope of the present study. What I focus on here is the features of the visual styles in Reading and Using Language within the macrogenre of the teaching unit The Olympic Games, with the purpose of explaining the underlying reasons for the different choices of visual styles. This is important because both styles target the same group of viewers.

6.6.2 Photograph for representing actual scene

The Reading section considered here (see Example (13) below) is a description of modern Olympic Games, with an emphasis placed on the similarities and differences
between the ancient and modern Olympics. The whole section spreads across two pages. The original layout and typeface can be found in Appendix III.

(13)

AN INTERVIEW

Pausanias, who was a Greek writer 2,000 years ago, has come on a magical journey to find out about the present day Olympic Games. He is now interviewing Lili, a Chinese girl.

P: My name is Pausanias, I lived in what you call “Ancient Greece” and I used to write about the Olympic Games more than 2,000 years ago. I have come to your time to find out about the present day Olympic Games because I know that in 2004 they are to be held in my hometown of Athens. Please can I ask you some questions?

L: Of course you can. What would you like to know?

P: How often do you hold your Games?

L: Every four years athletes from all over the world are admitted as competitors. There are two sets of Games – the Summer and the Winter Olympics and both are held every four years. The Winter Olympics are usually held two years before the Summer Olympics.

P: Winter Games? How can the runners enjoy competing in winter? And what about the horses?

L: Oh no! No running races or horse riding are included. There are events like skiing and ice skating which need snow and ice. That is why they are called the Winter Olympics.

P: Athletes competing from all over the world? Do you mean the Greek world? Our Greek cities used to compete against each other for the honour of winning. No other country could join in, nor could slaves or women.

L: All countries can take part if their athletes reach the standard to be admitted to the games. There are over 250 sports and each one has its own standard. Women are not only allowed to join in but play a very important role, especially in …

P: Please stop! All those events, all those countries and even women taking part! Where will all the competitors be staying?
L: A special village is built for the competitors to live in, a stadium for competitions, a very large swimming pool, a gymnasium as well as seats for those who watch the games.
P: It must be expensive. Does anyone want to host the Olympic Games?
L: As a matter of fact, everyone wants to. It’s a great honour. It’s just as much a competition among countries to host the Olympics as to win an Olympic medal. The 2008 Olympics will be held in China. Did you know that?
P: Oh yes! You must be very proud. Did you say medals? So even the olive wreath has been replaced! Oh dear! Do you compete for money too?
L: No, we don’t. It’s still all about being able to run faster, jump higher and throw further.
P: That’s good news! Thank you for your time. Goodbye.

Description is a kind of factual genre distinct from report, in the sense that description centres on specific facts about things while report is concerned with generic facts (Martin and Rose 2008: 5-6). In Example (13), the description is mainly given by a Chinese girl named Lili, who is interviewed by Pausanias, an ancient Greek who had written about the ancient Olympics and is now assumed to take a magical journey to the present day with the purpose of figuring out the modern Olympic Games. The specific description about the modern as well as ancient Olympics is as follows:

**Frequency of modern Olympics:**
every four years athletes from all over the world are admitted as competitors
the Summer and the Winter Olympics… both are held every four years
the Winter Olympics are usually held two years before the Summer Olympics

**Sports events in Winter Olympics:**
skiing and ice skating which need snow and ice

**Participants of ancient Olympics:**
Greek cities used to compete against each other
no other country could join in
slaves or women (could not join in)

**Scale of modern Olympics:**
all countries can take part if their athletes reach the standard to be admitted to the games
there are over 250 sports and each one has its own standard
women are not only allowed to join in but play a very important role
a special village is built for the competitors to live in, a stadium for competitions, a very large swimming pool, a gymnasium as well as seats for those who watch the games

**Host of modern Olympics:**  
a competition among countries to host the Olympics

**Award of modern Olympics:**  
Olympic medals  
still all about being able to run faster, jump higher and throw further

**Award of ancient Olympics:**  
the olive wreath

As was mentioned above, the description is embedded in an interview. *Pausanias* interrupts *Lili* everytime he is shocked by the great changes that have taken place in the modern Olympics, and *Lili* in return explains what has been replaced and clarifies what remains. Accordingly, a description of the major characteristics of modern Olympic Games is presented, as compared with those of the ancient one. There are three images accompanying this description: a photograph of the sculpture representing *Pausanias*’ head and shoulders, a photograph showing a scene of the opening ceremony of the modern Olympic Games, and a photograph that captures a moment of the speed skating competition in modern Olympics. These three photographs comply with the naturalistic standard that is defined by the 35mm photographic technology and represent the scenes in real life that we may see with the naked eye. Although the photograph of *Pausanias*’ sculpture might be seen as a borderline case due to its plain background, its naturalistic way of representing the colour, texture, and the play of light and shade convinces us of its naturalistic coding orientation.

It can be inferred that the choice of naturalistic visual style in Example (13) reflects the influence of the field of the discourse, i.e. the topic of the description. Although the interview is an imaginary one, *Pausanias* was a real writer in Greek history, as is revealed in the written text by the realizations of the noun *a Greek writer* and the accurate indication of circumstance *2,000 years ago*. The corresponding visual image represents the naturalistic sculpture of *Pausanias*, which is the evidence of his former presence that can be traced in our modern times. Here *Pausanias* is
represented as a historical figure rather than an imaginary comic style character, which helps focus the reader’s attention on the factual account of the similarities and differences between the ancient and modern Olympics instead of on the fanciful story about the miraculous journey. *Lili* is a fictitious character, and yet what she describes about the modern Olympic Games are facts that happen in real life. Therefore, none of the two images that accompany the interview and snapshot the modern Olympics is about *Lili* (which would require a comic style). Both photographs represent the real scenes of modern Olympic Games, which form a cohesion together with the nouns in the verbal text such as *the Summer Olympics, the Winter Olympics*, and *ice skating*. This choice of visual style again attaches great importance to the factual information about the modern Olympics that is given in the description rather than the imaginarily timeless interview between *Pausanias* and *Lili*.

To summarize, the real, factual orientation of the topic, or put it more technically, the field of the whole discourse determines the choice of the naturalistic visual style. The corresponding images that uphold naturalistic coding orientation, on the other hand, enhance the factual positioning of the discourse that the intended reader is supposed to assume.

### 6.6.3 Drawing for illustrating mythology

Another constituent genre of the macrogenre of the teaching unit is a mythic narrative entitled *The Story of Atlanta*, which tells the story about *Atlanta* and *Hippomenes*, two Greek mythological figures. The narrative is divided into two parts in the teaching section of *Using language*. The first part of the story is presented as a reading material, while the second part is recorded as listening material. The two parts are combined and the whole story is represented in Example (14) below.
THE STORY OF ATLANTA

Atlanta was a Greek princess. She was very beautiful and could run faster than any man in Greece. But she was not allowed to run in the Olympic Games. She was so angry that she said to her father she would not marry anyone who could not run faster than her. Her father said she must marry and asked her if she wanted to marry a king or prince. But Atlanta replied, “I will only be married to a man who can run faster than me. When a man says he wants to marry me, I will run against him. If he cannot run as fast as me, he will be killed. No one will be pardoned.”

Many kings and princes wanted to marry Atlanta but when they heard of her rules, some of them sadly went home. Other men stayed to run the race. There was a man called Hippomenes who was amazed when he heard of Atlanta’s rules. “Why are these men so foolish? Why will they let themselves be killed because they cannot run as fast as this princess?” Then when he saw Atlanta come out of her house to run, Hippomenes changed his mind, “I will marry Atlanta – or die!” he said.

The race started and although the men ran very fast, Atlanta ran faster. As Hippomenes watched, he thought, “How can I run as fast as Atlanta?” He went to ask for help from the Greek Goddess of Love. She promised to help him and gave him three golden apples. She said, “Throw an apple in front of Atlanta when she is running past and she will be relaxed. When she stops to pick it up, you will be able to run on and win.”

Hippomenes took the apples and went to the King. He said, “I want to marry Atlanta.” The King was sad to see another man die, but Hippomenes said, “I will marry her – or die!” So the race began.

LISTENING TEXT

THE RACE

So the next day the race began. Hippomenes was standing waiting and when Atlanta came out she thought, “I do not want this man to die. His death will not be caused by me!” So she said to her father, “Tell him to go away. The race will not be run today.” But Hippomenes said, “She has said it: she will be married to the man who runs faster than her. Come on now! Let’s run!”

Atlanta ran and Hippomenes ran too. He ran very fast but even so Atlanta ran faster. Soon she was in front of him. Then Hippomenes threw one of the golden apples. It went over her head and fell to the side of Atlanta. She stopped and had to run to get it. Hippomenes ran by. Atlanta looked up and she saw him in front. So she ran faster than
the fastest bird. She came near! Then she flew past him and was in front again. Again he threw another apple over her head and to the side. Again she stopped to pick it up.

When she reached Hippomenes the third time, he threw the third apple so far to the side that she had further to go. She saw it and wanted it. So she ran and picked it up. By that time Hippomenes was tired and could not run so fast. He was not so far in front. Atlanta ran fast but the apples were heavy and she could not catch up with him. So Hippomenes won the race and married Atlanta. Thanks to the Goddess they lived happily ever after.

Labov and Waletsky (1967/1997) propose a schema that consists of the stages of Orientation, Complication, Evaluation, Resolution and Coda to refer to the generalized structure for narratives of personal experience, with Complication and Resolution as obligatory stages while the others optional. In order to overcome the formalist bias that tends to subordinate Evaluation as the boundary between Complication and Resolution, systemicists have taken Labov and Waletsky’s narrative analysis as the point of departure and carried out a more discourse-oriented approach to the analysis of story types (Martin and Plum 1997; see also Martin and Rose 2008: 49-90). Martin and Rose (2008: 51-52) identify recount, anecdote, exemplum, and observation along with narrative in the story genre family. Narrative is specifically referred to ‘the generic pattern that resolves a complication’, and the evaluation of narrative complications allows variations between AFFECT, JUDGEMENT, and APPRECIATION, which are often employed to ‘suspend the action, increasing the narrative tension, and so intensifying the release when tension is resolved’ (Martin and Rose 2008: 52). The narrative structure of The Story of Atlanta is analyzed in Table 6-2 below, in which generic stages and phases within each stage (Martin and Rose 2008: 79-89) are specified.

### Table 6-2 The narrative structure of The Story of Atlanta

<table>
<thead>
<tr>
<th>Orientation</th>
<th>The Story of Atlanta was a Greek princess. She was very beautiful and could run faster than any man in Greece.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Setting problem</td>
<td>But she was not allowed to run in the Olympic Games.</td>
</tr>
</tbody>
</table>
She was so angry that she said to her father she would not marry anyone who could not run faster than her.

"I will only be married to a man who can run faster than me. When a man says he wants to marry me, I will run against him. If he cannot run as fast as me, he will be killed. No one will be pardoned."

Many kings and princes wanted to marry Atlanta but when they heard of her rules, some of them sadly went home. Other men stayed to run the race.

There was a man called Hippomenes who was amazed when he heard of Atlanta’s rules. “Why are these men so foolish? Why will they let themselves be killed because they cannot run as fast as this princess?” Then when he saw Atlanta come out of her house to run, Hippomenes changed his mind, “I will marry Atlanta – or die!” he said.

The race started and although the men ran very fast, Atlanta ran faster.

As Hippomenes watched, he thought, “How can I run as fast as Atlanta?” He went to ask for help from the Greek Goddess of Love. She promised to help him and gave him three golden apples. She said, “Throw an apple in front of Atlanta when she is running past and she will be relaxed. When she stops to pick it up, you will be able to run on and win.”

Hippomenes took the apples and went to the King. He said, “I want to marry Atlanta.” The King was sad to see another man die, but Hippomenes said, “I will marry her – or die!” So the race began. (So the next day the race began.)

Hippomenes was standing waiting and when Atlanta came out she thought, “I do not want this man to die. His death will not be caused by me!”

So she said to her father, “Tell him to go away. The race will not be run today.”

But Hippomenes said, “She has said it: she will be married to the man who runs faster than her. Come on now! Let’s run!”

Atlanta ran and Hippomenes ran too. He ran very fast but even so Atlanta ran faster. Soon she was in front of him. Then Hippomenes threw one of the golden apples. It went over her head and fell to the side of Atlanta. She stopped and had to run to get it. Hippomenes ran by.
problem Atlanta looked up and she saw him in front. So she ran faster than the fastest bird. She came near! Then she flew past him and was in front again.
solution Again he threw another apple over her head and to the side. Again she stopped to pick it up. When she reached Hippomenes the third time, he threw the third apple so far to the side that she had further to go. She saw it and wanted it. So she ran and picked it up.
problem By that time Hippomenes was tired and could not run so fast. He was not so far in front.
solution Atlanta ran fast but the apples were heavy and she could not catch up with him. So Hippomenes won the race and married Atlanta.

Coda

comment Thanks to the Goddess they lived happily ever after.

In the beginning the Orientation (i.e. from Atlanta was a Greek princess to Other men stayed to run the race) presents the heroine Atlanta’s identity (i.e. a Greek princess) and the location of the story (i.e. in Greece), her speciality (i.e. very beautiful and could run faster than any man), and the rule that gives rise to a problem (i.e. not allowed to run in the Olympic Games). It is followed by Atlanta’s reaction to the rule (i.e. so angry), and another problem for Atlanta (i.e. Her father said she must marry and asked her if she wanted to marry a king or prince). The counterexpectancy signalled by ‘but’ creates tension and leads to another problem for suitors, i.e. Atlanta’s insistence that any man who wants to marry her must run faster than her or will be killed, preceding the suitors’ different reactions. The Complication (i.e. from There was a man called Hippomenes who was amazed when he heard of Atlanta’s rules to Soon she was in front of him) describes how Hippomenes marries Atlanta. The hero Hippomenes is introduced with his reflection on how to win the race. It then describes his seeking advice from the Greek Goddess of Love and his marriage proposal. Though Atlanta does not want Hippomenes to die because of losing the race, Hippomenes’ insistence creates tension again. The Resolution (i.e. from Then Hippomenes threw one of the golden apples to So Hippomenes won the race and married Atlanta) clarifies Hippomenes’ success with the help of the three golden apples given by the Greek Goddess of Love. Everytime problem arises and Hippomenes almost loses the race, he throws an apple, which
arouses Atlanta’s desire and the tension is resolved. Hipponemes eventually wins the race, and the Coda (i.e. *Thanks to the Goddess they lived happily ever after*) presents the narrator’s comment on the happy ending.

There are two images in *The Story of Atlanta*, one depicting Atlanta’s negotiation with her father about her marriage and the other illustrating Hipponemes receiving the three golden apples from the Greek Goddess of Love. Both images are colour drawings. From the point of view of the naturalistic standard, they are of high modality in terms of colour saturation, colour differentiation, colour modulation, representation, and brightness. For instance, efforts are made to depict the skin colour, the textures of people’s hair and clothes in a subtle way and as relatively close to what we see with the naked eye. Nevertheless, the images depart from the naturalistic coding orientation in the aspects of illumination, depth, and contextualization. To be specific, there is no play of light and shade in the images, and the simple overlapping leads to the absence of depth. In terms of contextualization, the decoration of the palace and the scenery outside the palace that serve as the background in the first image is selectively simplified, and what are reserved are confined to those architectural features that represent Greek characteristics. As for the background in the second image, the drifting clouds around the lady enhance the indication of her identity as the Goddess.

The targeted viewers of the aforementioned two images are senior secondary students, and naturalistic coding orientation is retained in the representations of certain modality markers as have been analyzed above. This complies with the general truth criteria in the whole senior secondary EFL textbooks. However, the deviations from this naturalistic standard are observed in illumination, depth, and contextualization. I attempt to account for this by resorting to the register variable of field. As was previously discussed, the topic of the whole discourse is a Greek mythology that is related to the Olympic Games. The mythology is supposed to be helpful in students’ understanding of the culture and traditions of ancient Greece in which the Olympics is developed, and yet the mythology itself and those characters involved in it may not be historical event or figures that have ever existed.
Moreover, the plot of the mythology is more imaginary than factual. Therefore, it is drawings rather than photographs that are employed to illustrate the remote, ancient story whose historical facticity is hard to nail down. The choice of this fanciful style, on the other hand, induces the putative reader to agree on the imaginary characteristics of the mythic narrative.

From the discussion of the two images within a macrogenre, it can be inferred that the register variable of field is one of the important factors that exerts influence on the choice of visual style. Certain modality markers may be adjusted to some extent due to the topic or focus of the social activity, and the relevant coding orientation will be modified accordingly, thus resulting in the style variations like photographs and drawings that target the same group of viewers.

6.7 Summary

The present chapter has focused on the issue of truth criteria in the multimodal EFL textbook discourse. After a brief retrospect of relevant studies on modality in both monomodal and multimodal discourse, I have first explicated with exempla from the data the visual parameters whose configurations express modality of a given visual event and the coding orientations that provide criteria for modality judgement in different contexts. A comparative analysis has then been conducted to examine coding orientation in three teaching units for primary, junior and senior secondary EFL education, so as to investigate the ‘interdependence’ (Hodge and Kress 1988: 161) between modality and social relations from a multilevel perspective in the context of EFL education. Furthermore, the childist ideology that underlies the overdose of cartooning in primary and junior secondary EFL textbooks has been discussed in relation to a sociological position towards pedagogic discourse and nature of knowledge. Before winding down the chapter, I have also considered the influence of field on the choice of visual style within the same teaching unit as a macrogenre.

The analysis has shown that what counts as real is culturally defined and specific
to a given communicative context or among members of a certain social group. In the primary EFL textbook discourse, the pictorial techniques used in cartoons follow the pleasure principle, and hence sensory coding orientation is adopted. However, it is argued that the pervasive use of cartoons whose main purpose is purely for fun in educational setting needs to be rethought. The sense of abstractness mounts upwardly as students move through the school years from primary to junior secondary education. On the one hand, the vibrant, exciting colouration still acts as a source of pleasure, while on the other, the abstract representation and contextualization of the images capture the underlying essence of what is represented. Accordingly, it can be inferred that abstract-sensory coding orientation is adopted in junior secondary EFL textbook discourse. Last but not least, the photographs in senior secondary EFL textbook adopt naturalistic coding orientation in the teaching unit with the similar topic on animals, which is the shared, dominant truth criteria in modern society as is defined by the current photographic technology. In this context the viewers are treated as members of the general public who share the dominant cultural norms. Based on the analysis and discussion above, it is justified to say that the resources that convey modality value in EFL textbook discourse are multimodal, and the way these multimodal meaning-making resources are deployed both reflects and constructs the solidarity relationship between textbook editors and different groups of English learners. Nevertheless, within the educational setting how much concession the textbook editor should make to cater for the alleged ‘child-centreness’ without sacrificing the pedagogic purpose warrants scholarly attention and needs further exploration.

As have been observed in the data, visual styles with different degrees of deviation from the accepted truth criterion can be employed in different constituent genres of the macrogenre of a given text. The remainder of this chapter has demonstrated that the choice of visual style is influenced to some extent by the topic or focus of the social activity, and the visual style on the other hand enhances the putative positioning that the intended reader is supposed to assume.
NOTES

1 There is application of colour in visual displays in certain scientific and technological texts. Nevertheless, this use of colour is ‘more or less arbitrary to facilitate the reading of complex diagrams’ rather than the actual colour that is encountered in reality (Kress and van Leeuwen 2006: 169).

2 Further studies in this area have been carried out in relation to genre-based literary pedagogy (e.g. Martin 1999b).

3 As Christie (2004: 174) explains, ‘authority’ here is in the sense of ‘authoritative’ rather than ‘authoritarian’.

4 For overviews and further discussion of Bernsteinian theory on pedagogy and knowledge see Bernstein 1990, 1999, 2000; Christie and Martin 2007).

5 According to Hyon (1996), within linguistic studies two other main perspectives on genre are English for specific purposes (ESP) and North American New Rhetoric studies.

6 Within SFL there are alternative viewpoints on the status of genre. Halliday (1978: 145) treats genre as ‘an aspect of what we here call the “mode”’. As for the concept of generic structure, he articulates that it ‘can be brought within the general framework of the concept of register, the semantic patterning that is characteristically associated with the “context of situation” of a text’(Halliday 1978: 134). Hasan (1985: 63-69) emphasizes the relationship between text structure, genre and field. These perspectives tend to relate genre to one of the contextual variables, rather than viewing genre as a more abstract semiotic stratum that is above register.

7 When analyzing generic structure of story, Martin and Rose (2007b) suggest adopting both perspectives of generic stages and phases. Story phases have to do with ‘a common set of resources’ that the story family (i.e. narrative, recount, anecdote, exemplum, and observation) share in ‘moving sequences forward and engaging readers’ (Martin and Rose 2007b: 79). Each phase type has a specific function and may occur in any stage of a story.

8 I am grateful for Dr. David Rose’s comments on the phase analysis of The Story of Atlanta, though all mistakes are mine.
Chapter 7

Conclusions and Suggestions

7.1 Introduction
The overall aim of this thesis has been to account for the multimodal construal of interpersonal meaning in China’s EFL context, and I have concentrated the discussion on the multimodal instances in EFL textbooks for primary and secondary education. In order to understand the ways in which meanings are created visually as well as verbally, I have drawn upon APPRAISAL theory, a model developed within SFL for analyzing interpersonal meaning at the level of discourse semantics, and the systemic functional semiotic approach when examining verbal and visual semiotic systems. This chapter is intended to bring together the major findings of the three sub-studies reported in the previous chapters, and to discuss the implications of the work in terms of its theoretical as well as pedagogic contributions. As a final point, the thesis will conclude with the identification of some areas that require further research.

7.2 A retrospect of the research findings
As was mentioned in the introductory chapter, the current research grows out of the newly-emerged research frontier of multimodality. Textbooks for China’s EFL context, which have been less frequently studied as compared with multimodal pedagogic materials for other school subjects, are chosen as the focus of the present study. The systemic functional semiotic approach, which provides theoretical foundations on which the current work is grounded, has proven to be effective in
dealing with texts that involve more than one semiotic system. In response to the research aims set out at the beginning of the thesis, I shall summarize the key research findings of the previous chapters. In Chapter 1 the following research questions are formulated to specify the research focus:

1. How are linguistic and visual meaning-making resources deployed to mediate the heteroglossic space in the multi-voiced EFL textbook discourse? In what way the degree of alignment between voices may be graded by utilizing multimodal resources?

2. How is the evaluative stance in multimodal EFL textbook discourse construed through linguistic and visual semiotic systems? What kinds of verbiage-image relations are there in the intermodal construal of attitudinal meanings?

3. In what way are multimodal messages defined as real in the EFL textbooks targeting readers of different educational levels? How is it conditioned by as well as construing the context in which it is embedded? Is the visual arrangement in line with the nature of pedagogic discourse, and why?

Analyses in Chapters 4 to 6 have unfolded to encompass three sub-studies, i.e. ‘heteroglossic harmony’, ‘accumulative attitude’, and ‘contestable reality’, to tackle three different facets of interpersonal management in EFL textbook discourse. As will be explained below, the aims set out in Chapter 1 have been properly achieved.

7.2.1 Recognition of multimodal ENGAGEMENT devices and their gradability

To answer the first research question, I have identified and examined in Chapter 4 the multimodal meaning-making resources that enable dialogic engagement in EFL textbook discourse, and discussed how these multimodal resources scale up or down engagement values. It is found that along with the authorial editor voice, a given proposition or evaluation can be attributed to other sources, i.e. character voice and reader voice, through the use of multimodal ENGAGEMENT devices. Analyses of five types of ENGAGEMENT devices (i.e. labelling, dialogue balloon, jointly-constructed
text, illustration and highlighting) reveal that they realize the ENGAGEMENT meanings of [disclaim], [proclaim], [entertain] and [attribute] (Martin and White 2005: 97-98). Specifically, labelling enables editor voice to negotiate meanings with character voice by fending off alternative positions, hence insisting upon the prescribed teaching goals and realizing [proclaim]. Dialogue balloon and illustration bring in character voice, and thus the proposition or viewpoint is “attributed” to character(s). Jointly-constructed text opens up space to introduce reader voice, encoding [attribute] as well. Although the editor-reader alignment does not mean a total agreement between the communicative parties, jointly-constructed text basically construes a compliant reading position, in that its completion requires some shared knowledge and cooperative performance from the ideal reader. Within certain illustrations, editor voice and character voice may go against one another and hence [disclaim] is realized. Highlighting in visual image may function to [entertain] other possibilities that are grounded within the contingent subjectivity of the multimodal text itself.

The gradability of ENGAGEMENT values (Martin and White 2005: 135) realized by the multimodal resources has also been considered. It is argued that dialogue balloon may realize different degrees of [attribute] values based on the amount (i.e. FORCE) of responsibility that characters undertake in instructing the reader. Different degrees of reader involvement are closely related to the degree of completion or fulfillment (Hood and Martin 2007) of a jointly-constructed text, i.e. its prototypicality (i.e. FOCUS) as a co-construction. The higher degree of completeness a jointly-constructed text possesses, the less heteroglossic space it opens up to reader voice. In terms of illustration, the intensity of [attribute] meaning can be approached by looking at the indispensability of the image in linking different components of a verbal text.

After the identification and analyses of the aforementioned five ENGAGEMENT devices, I further examined the interaction between multiple voices in EFL textbook discourse in relation to three parameters in visual interaction, i.e. contact, social distance and point of view. It is shown that while the majority of images belong to the ‘observe’ type, eye contact between character and viewer is established in those
images where reader’s action is required. The various choices of social distance demonstrate a wide range of social relations between characters and the reader. The frontal and eye-level angle is frequently adopted due to the engagement of characters that are represented as peers of the reader.

7.2.2 Intermodal construal of the ‘emotion and attitude’ goal
As was suggested in Chapter 3, the exploration of interpersonal meaning also calls for an examination of the ‘personal’ dimension. In an attempt to arrive at the answer to the second research question, Chapter 5 has been devoted to the analysis of the way that multimodal meaning-making resources are deployed to achieve the ‘emotion and attitude goal’ in EFL education (Cheng 2002; Zhong 2006; Zhu 2006), with particular attention paid to the intersemiotic relations, i.e. complementarity and co-instantiation in the construal of attitudinal meanings.

The various styles of visual images (i.e. cartoon, portrait, and photograph) in EFL textbooks are grouped into two categories according to whether they describe human participants or inanimate things and landscapes. It is found that the frequently-adopted cartoon style in primary and junior secondary textbooks is often indicative of positive affectual meanings, and the happiness of English-learning is thus directly inscribed. The corresponding verbiage, on the other hand, hardly commits any attitudinal meanings but simply informs the reader of the teaching content. The image-verbiage relation is thus complementary, in that the attitudinally hypo-committed verbal texts explain to the reader what is supposed to be learned, whereas cartoons visually construe an evaluative stance that is assumed to be taken during the process of learning. Some cartoons may also function to invoke APPRECIATION through logogenetic recontextualization (Martin 2002). It can be inferred from the analysis that the visual semiotic system bears much of the responsibility for encoding attitudinal meanings in the textbooks for primary and junior secondary schooling. The trend for explicit affectual inscriptions gradually declines in the textbooks for secondary education. In other words, AFFECT and FORCE are ‘coupled’ (Martin 2008b) in different ways at these two levels of education.
The emerging serious images signal a transition from the playful childhood to the adolescence, where a sense of social responsibility is to be cultivated. As for senior secondary textbooks, the serious portraits and photographs with implicit JUDGEMENT or APPRECIATION invocations (Martin and White 2005: 66-67) work in concert with the verbal attitudinal inscriptions and invocations. In this case, both language and images commit attitudinal meanings and thus the image-verbiage relation is co-instantiation. The two semiotic systems echo and reinforce one another in establishing the overall evaluative stance assumed in the putative readership. Image and verbiage also work in tandem with each other in guiding the reader in the completion of jointly-constructed texts. It is shown that an overall evaluative stance can be established in the verbiage, whereas the images visualize the ideational tokens as well as invoke positive APPRECIATION, which reinforces the verbal attitudinal orientation. In terms of the amount of meaning potential that is instantiated in different types of visual images, it is contended that indirect invocations commit less attitudinal meanings than overt inscriptions.

Based on the studies of the ontogenesis of ATTITUDE system, this sub-study has concluded by accounting for how attitudinal meanings are institutionalized and accumulated as the students progress all the way from primary to secondary education. It is argued that a gradual shift in ATTITUDE encoding can be identified – from a personal emotional release to a more institutionalized method of evaluation, which reflects as well as contributes to the accumulation of institutionalized feelings stipulated in the current curriculum standards.

7.2.3 Influence of register variables on coding orientation

While Chapter 5 deals with the attitudinal aspect of interpersonal meaning, the third sub-study in Chapter 6 focuses on the issue of modality (Hodge and Kress 1988; Kress and van Leeuwen 2006; van Leeuwen 2005), i.e. the question of ‘truth’ that a message producer wants a representation to induce in the intended receiver. One of the important tenets in SFL is the close relation between text and context, and the third research question asks about the influence of tenor and field on the criteria for
modality judgements in multimodal texts. Drawing on the eight modality markers and four types of coding orientation developed by Kress and van Leeuwen (2006), the third sub-study provides a multilevel, comparative analysis of coding orientation in three teaching units on a similar topic for different educational contexts.

It is found that the way in which multimodal resources are deployed to convey modality values in pedagogic context is both conditioned by and construing the solidarity between textbook editors and different groups of learners. Specifically, the cartoons in primary EFL textbooks follow the ‘pleasure principle’ and adopt a sensory coding orientation (Kress and van Leeuwen 2006). Though sensation remains in junior secondary textbooks, the sense of abstractness tends to mount upwardly along the educational ladder. Nevertheless, it is suggested that the amount of cartooning in the primary and junior secondary materials is excessive from the perspective of educational theory on pedagogic discourse and nature of knowledge (Bernstein 1990, 2000). The photographs in senior secondary textbooks, on the other hand, employ a naturalistic coding orientation, positioning the viewers as members of the general public who share the dominant cultural norms.

In addition to the multilevel view on modality in three different EFL educational contexts, the third sub-study also considers various degrees of deviation from the accepted truth criterion in multimodal texts that target the same group of viewers. It is argued that another register variable, i.e. field, may exert some influence on the choice of visual style in different constituent genres within the same teaching unit as a macrogenre. This choice of visual design, on the other hand, helps to enhance an intended reading positioning. The examination has shown that what counts as real in multimodal texts is culturally defined and specific to a given communicative context.

In sum, it can be inferred from the three sub-studies that multimodal meaning-making resources play a significant role in encoding interpersonal meaning. The visual semiotic system in multimodal EFL textbook discourse is by no means redundant with language in meaning-making. In terms of editor-reader alignment, linguistic and visual semiotic systems combined to enable the interpolation of character voice and offer diverse means for editor voice to engage the reader with the
proposition put forward in a text. As for the ‘emotion and attitude’ goal highlighted in the current *Curriculum Standards for English*, the intermodal construal contributes to the accumulative attitudinal meanings in primary and secondary education. Last but not least, the intended truth-value of a multimodal message is closely related to the socially defined truth criterion within a specific social group, though deviation is allowed to cater for different topics within a teaching unit. Nevertheless, it is argued that the nature of pedagogic discourse should also be taken into account when designing visual display for pedagogic materials. As was explained in Chapter 1, this thesis is theoretically as well as pedagogically motivated. In the ensuing section, I shall outline some contributions the present study has tried to make to the research on interpersonal semantics in multimodal texts, as well as its pedagogic implications for EFL education.

### 7.3 Contributions and implications

The theoretical and practical relevance of the current research is two-fold. One is to work towards a comprehensive model of APPRAISAL (Martin 2000a; Martin and White 2005) that accounts for both linguistic and visual semiotic systems, and the other is to address pedagogic concerns closely related to multimodality.

#### 7.3.1 Theoretical contributions

The APPRAISAL system (Martin 1997, 2000a; Martin and White 2005) is a functional model that accounts for interpersonal meaning at the level of discourse semantics. As was stated in Chapter 3, a review of the literature on APPRAISAL reveals that while the growing body of work in this area has covered visual semiotic system, most of the studies have been oriented to the sub-system of ATTITUDE (e.g. Economou 2006), leaving the aspects of ENGAGEMENT and GRADUATION remain under-theorized. This thesis has attempted to explore new ground by identifying and analyzing multimodal resources that may realize ENGAGEMENT values. It can be inferred from the findings that, on the one hand, one type of ENGAGEMENT meaning may be realized through
different multimodal devices (e.g. [Attribute] can be realized by dialogue balloon or illustration). On the other, one type of multimodal resource may encode different ENGAGEMENT meanings (e.g. Illustration may function to encode [attribute] or [disclaim]).

Another contribution of the present study to the understanding of multimodal APPRAISAL resources lies in the examination of the ways in which the attitudinal meanings (i.e. AFFECT, JUDGEMENT, and APPRECIATION) are realized through the intersemiotic co-instantiation and complementarity. It has been demonstrated that attitudinal meanings are realized in verbal texts and visual images via both explicit inscription and implicit invocation. In those multimodal texts where verbiage is attitudinally under-committed (see Martin 2008b, Hood in press for an introduction to the concept of ‘commitment’), visual images may step up to reinforce the evaluative stance. When image and verbiage work together in the construal of value positions, these two semiotic systems co-articulate with one another in establishing the attitudinal orientation.

A further theoretical contribution developed in this research is the exploration of how the options from ATTITUDE and ENGAGEMENT can be ‘coupled’ (Martin 2008b) with the options from GRADUATION (i.e. FORCE and FOCUS), and thus various degrees of attitudinal meanings and ENGAGEMENT values can be encoded in visual communication. It is indicated that different degrees of the FORCE of inscribed AFFECT can be realized through a variety of depictions of facial expressions and behavioural manifestations that imply different degrees of affectual intensity. It is also found that the ways in which ENGAGEMENT values can be scaled are strongly associated with the intrinsic property of a multimodal resource, such as the projective structure of dialogue balloon that mediates a character’s utterance, the co-constructedness of jointly-constructed text, or the elucidative and supportive role of illustration.

### 7.3.2 Pedagogic implications

Along with the contributions to the theoretical model that underpins the present
Conclusions and Suggestions

research, this thesis has set for itself a pedagogic goal that aims at providing some implications for EFL education and textbook design. As was articulated in the summarizing remarks of Chapter 2, fives gaps have been recognized in the previous pedagogic research. Each of the three sub-studies in the current research has attempted to suggest some answers to at least one aspect of the five areas. The investigation into the heteroglossic nature of EFL textbook discourse demonstrates from a semiotic perspective the way in which multiple semiotic resources are involved in the ‘dialogic process’ that is advocated in the current classroom teaching (e.g. Chen and Ye 2006), showing how multiple semiotic resources can be manipulated to bring in voices other than the editor’s and align the intended reader with propositions and viewpoints.

As has been reiterated in this thesis, the current curriculum reform in China points to the significance of emotions, attitudes, and values education (Zhu 2006). The current research has strived to anchor the discussion onto linguistic and semiotic analysis, which is intended to reveal the role that multiple semiotic systems can play in realizing the ‘emotion and attitude’ goal. From the perspective of genre-based literacy (e.g. Martin and Rose 2007b, 2008), critiques are made on the primary and junior secondary EFL textbooks for their neglect of generic/schematic structure in the verbal texts, which might risk reversing the course of natural language learning. While showcasing the importance of multimodal resources in managing heteroglossic space and encoding attitudinal meanings, the work undertaken here offers a reminder that language teaching should follow a top-down approach from the level of genre through discourse semantic patterns downwards to specific lexicogrammatical features.

Still another pedagogic implication the current research arrives at is concerned with the choice of coding orientation in relation to the register variables of field and tenor. There is an evident association between image style and social relations. Nevertheless, it is argued that textbook editors and designers should bear in mind the fact that EFL textbooks is a kind of pedagogic discourse when designing visual arrangement for textbook discourse. It is hoped that with the affordances as well as
Conclusions and Suggestions

limitations of multiple semiotic systems made explicit from a social semiotic perspective, we may have one step further towards a comprehensive and critical understanding of multimodal EFL pedagogic materials.

7.4 Suggestions for future studies

The work undertaken in this thesis is one of the first efforts to probe into the multimodal nature of EFL pedagogic materials in China from a systemic functional semiotic perspective. This section is intended to give an indication of some pedagogic and theoretical aspects that may need further exploration.

7.4.1 Multimodality research in educational context

While the focus of this thesis is on primary and secondary textbooks, it may point to a similar examination of the multimodal nature of tertiary textbooks. Furthermore, within pedagogic context there are undoubtedly other types of multimodal communication outside textbook discourse. The multimodal teaching and learning resources widely used in China’s educational context, such as pictures, maps, three-dimensional objects, audio-visual devices, demand the attention from social semiotic analysts. In classroom practice, teacher’s gestures may be analyzed and discussed from a functional and semiotic perspective (c.f. Hood 2007). With the advances in communications technology, electronic and multimedia resources for pedagogic purpose are increasing in type and number. CD-ROMs and on-line resources used in computer-aided instruction also call for further theoretical development in the description of hypertext (c.f. Djonov 2007; Knox 2007). All these commonly found pedagogic materials that involve more than one semiotic system open up large arena for future multimodality research in educational setting.

7.4.2 Other aspects of a metafunctionally diversified view

Theoretically speaking, the adoption of the metafunctional principle within SFL reminds us that there are two other functional diversifications to be explored apart
Conclusions and Suggestions

from the interpersonal metafunction under attention. Ideationally, the process types involved in narrative representation or conceptual representation (Kress and van Leeuwen 2006: 45-113) can be further analyzed when examining multimodal pedagogic materials. The depiction of circumstantial elements such as setting is also worth investigating. Moreover, a discourse semantic perspective could be adopted by looking at the ideational meanings in visual media in terms of the relations between elements (i.e. ‘taxonomic relations’), the configuration of elements (i.e. ‘nuclear relations’), and the relations between processes (i.e. ‘activity sequence’) (Martin and Rose 2007a: 76). Textually or compositionally, modern printing and digital technologies make possible a great diversity of layout designs. Systemic functional semiotic approach may provide an effective analytical tool when addressing issues that arise from various visual composition (e.g. horizontal polarization, vertical polarization, triptych, etc; see Kress and van Leeuwen 2006). Of course, the possible realms mentioned here are merely part of the picture, which may give us a glimpse of the great potential of social semiotic theory in the examination and explanation of multimodal texts.

7.4.3 Inspiration from other contexts

Looking beyond the pedagogic setting, researchers may find that other communicative contexts can provide testing grounds for the expansion and elaboration of social semiotic theory. Take the elaboration of ENGAGEMENT and GRADUATION systems in the current research for instance. It is identified in Chapter 4 that there are five types of multimodal resources serving as ENGAGEMENT devices (i.e. labelling, dialogue balloon, jointly-constructed text, illustration and highlighting). These findings are based on the examination of EFL textbook discourse. Other types of multimodal meaning-making resources that encode ENGAGEMENT meanings may be recognized in analyzing multimodal texts in other contexts. With the identification of more ENGAGEMENT devices and the specification of their different functions, the GRADUATION network for the scaling of values in multimodal discourse may accordingly be established and expanded. Addressing these theoretically oriented
issues calls for considerable data and in-depth discussions, while at the same time will bring us closer to a comprehensive understanding of linguistic and semiotic theories.

7.5 Envoi

The research reported in this thesis has endeavoured to bring a knowledge of linguistic and visual semiotic systems to an understanding of multimodal pedagogic materials for teaching English as a foreign language. Both multimodality and English language teaching warrant close attention from linguists. Throughout the thesis I have sought, albeit briefly, to systematically consider visual semiotic system along with language, which will hopefully offer insights into the research on either of them or both. The site of EFL education, as has been shown in this work, proves to be one of the major contexts for a close examination of multiple semiotic systems. The exploration of multimodal construal of meanings, on the other hand, may address some of the practical issues arising from the current attitudes towards and demands for English teaching. While the focus of the present study is mainly on interpersonal meaning, it is predictable that multimodality and pedagogic materials will continue to constitute part of the ongoing linguistic and semiotic agenda in the foreseeable future.
References


Bo, T. (博涛) 2000. Qianxi dili jiaocai zhong ditu de sheji sixiang (浅析地理教材中地图的设计思想, ‘Analysis of the map design in geography textbooks’).

Kecheng jiaocai jiaofa (课程 • 教材 • 教法, ‘Curriculum, Teaching Material and Method’) (12): 40-43.


Djonov, E. 2007. Website hierarchy and the interaction between content organization, webpage and navigation design: A systemic functional hypermedia discourse
References


Fu, Y. (傅莹) 2005. Yingxiang de jiaoyu liliang (影像的教育力量, ‘The educational
References

power of visual images’). Guangdong jiaoyu (广东教育，‘Guangdong Education’), (7): 60-61.


References

Oxford University Press.


Iedema, R. 1995. Literacy of Administration (Write it Right Literacy in Industry Research Project - Stage 3). Sydney: Metropolitan East Disadvantaged Schools Program.


References

*Literacy in First and Second Languages: Meaning with Power.* London: Lawrence Erlbaum Associates, 21-44.


Society and Culture (Sydney Studies in Society and Culture), 225-268.


Martin, J.R. and White, P.R.R. 2005. The Language of Evaluation: Appraisal in
Mukarovsky, J. 1978. *Structure, Sign and Function*. (Burbank J. and Steiner, P. trans.}


References

*PEP Primary English Students’ Book II for Year 5.* 2003. Beijing: People’s Education Press.


Scollon, R. 1997. Handbills, tissues, and condoms: A site of engagement for the


References

*Contexts of Text and Image in Classroom Practice*. Buckingham: Open University Press.


References


Xie, H.J. and Song, N.Q. (谢华均，宋乃庆) 2003. Xin jiaocai bianxie de jiaoyu linian tanxi (新教材编写的教育理念探析，‘Exploration of the educational principles underlying the compilation of the new textbook series’).


Ye, Q.C. (叶起昌) 2006. Chaowenben duoyushi de shehui fuhaoxue fenxi (超文本多语式的社会符号学分析，‘A social semiotic analysis of hypertext’s multimodality’).


Net.1. General Semiotics (from ‘J.L.Lemke Online Office’)

http://www-personal.umich.edu/~jaylemke/theories.htm
Appendix I  A multimodal page introducing major characters in primary textbooks

[PEP Primary English Students' Book I for Year 6]
Appendix II The exact layout of *SIGHTSEEING IN LONDON*

5 On my way to the station my car broke down. When I got to the repair shop I _______ (it / close).

6 The computer doesn’t seem to work well. You’d better _______ (it / repair)?

7 Jill and Eric _______ (all their money / steal) while they were on holiday.

8 Chris _______ (some flowers / send) to Sarah on her birthday. Then Chris asked Sarah to marry him and they _______ (it / announce) in the newspaper.

They had no time to arrange their own wedding, so they _______ (it / organize) by a company.

3. Play the game: “What did they find?” Write one sentence about what Mr and Mrs Smith found when they came home from work one day. Remember to use the past participle as the object complement.

**EXAMPLE:** They found the window broken.

Divide your class into groups of five. Give a piece of paper to the first person in each group who folds the paper over his / her writing, so the next person cannot see it. Pass the paper to the next person in the group. When the last one in the group has finished writing, he / she comes to the front of the class to read all the ideas of his / her group. The one with the most interesting ideas wins.

---

Using Language  

**SIGHTSEEING IN LONDON**

Worried about the time available, Zhang Pingyu had made a list of the sites she wanted to see in London. Her first delight was going to the Tower. It was built long ago by the Norman invaders of AD 1066. Fancy! This solid, stone, square tower had remained standing for one thousand years. Although the buildings had expanded around it, it remained part of a royal palace and prison combined. To her great surprise, Zhang Pingyu found the Queen’s jewels guarded by special royal soldiers who, on special occasions, still wore the four-hundred-year-old uniform of the time of Queen Elizabeth I.

There followed St Paul’s Cathedral built after the terrible fire of London in 1666. It looked splendid when first built! Westminster Abbey, too, was very interesting. It contained statues in memory of dead poets and writers, such as Shakespeare. Then just as she came out of the abbey, Pingyu heard the famous sound of the clock, Big Ben, ringing out the hour. She finished the day by looking at the outside of Buckingham Palace, the Queen’s house in London. Oh, she had so much to tell her friends!
The second day the girl visited Greenwich and saw its old ships and famous clock that sets the world time. What interested her most was the longitude line. It is an imaginary line dividing the eastern and western halves of the world and is very useful for navigation. It passes through Greenwich, so Pingyu had a photo taken standing on either side of the line.

The last day she visited Karl Marx's statue in Highgate Cemetery. It seemed strange that the man who had developed communism should have lived and died in London. Not only that, but he had worked in the famous reading room of the Library of the British Museum. Sadly the library had moved from its original place into another building and the old reading room was gone. But she was thrilled by so many wonderful treasures from different cultures displayed in the museum. When she saw many visitors enjoying looking at the beautiful old Chinese pots and other objects on show, she felt very proud of her country.

The next day Pingyu was leaving London for Windsor Castle. "Perhaps I will see the Queen?" she wondered as she fell asleep.

Make a list of Zhang Pingyu's tour of London and a comment on each place she visited.

<table>
<thead>
<tr>
<th>Day 1</th>
<th>Comments</th>
<th>Day 2 and comment</th>
<th>Day 3 and comment</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>4</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Appendix III  The original layout and typeface of *AN INTERVIEW*

Pausanias, who was a Greek writer 2,000 years ago, has come on a magical journey to find out about the present day Olympic Games. He is now interviewing Lili, a Chinese girl.

P: My name is Pausanias. I lived in what you call “Ancient Greece” and I used to write about the Olympic Games more than 2,000 years ago. I have come to your time to find out about the present day Olympic Games because I know that in 2004 they are to be held in my hometown of Athens. Please can I ask you some questions?

L: Of course you can. What would you like to know?

P: How often do you hold your Games?

L: Every four years athletes from all over the world are admitted as competitors.

There are two sets of Games – the Summer and the Winter Olympics and both are held every four years. The Winter Olympics are usually held two years before the Summer Olympics.

P: Winter Games? How can the runners enjoy competing in winter? And what about the horses?

L: Oh no! No running races or horse riding are included. There are events like skiing and ice skating which need snow and ice. That is why they are called the Winter Olympics.

P: Athletes competing from all over the world? Do you mean the Greek world? Our Greek cities used to compete against each other for the honour of winning. No other country could join in, nor could slaves or women.

L: All countries can take part if their athletes reach the standard to be admitted to the games. There are over 250 sports and each one has its own standard. Women are not only allowed to join in but play a very important role, especially in ...

P: Please stop! All those events, all those countries and even women taking part! Where will all the competitors be staying?

L: A special village is built for the competitors to live in, a stadium for competitions, a very large swimming pool, a gymnasium as well as seats for those who watch the games.

P: It must be expensive. Does anyone want to host the Olympic Games?

L: As a matter of fact, everyone wants to. It’s a great honour. It’s just as much a competition among countries to host the Olympics as to win an Olympic medal. The 2008 Olympics will be held in China. Did you know that?

P: Oh yes! You must be very proud. Did you say medals? So even the olive wreath has been replaced! Oh dear! Do you compete for money too?

L: No, we don’t. It’s still all about being able to run faster, jump higher and throw further.

P: That’s good news! Thank you for your time. Goodbye.