

# **A Picture is Worth a Thousand Words – or Is It?**

## **The Interplay of Text and Images in Technical Documents**

Aliisa Mäkynen  
University of Tampere  
School of Language,  
Translation and Literary Studies  
English philology  
Pro Gradu Thesis  
April 2012

Tampereen yliopisto  
Englantilainen filologia  
Kieli-, käännös- ja kirjallisuustieteiden yksikkö

MÄKYNEN, ALIISA: A Picture is Worth a Thousand Words – or Is It?  
The Interplay of Text and Images in Technical Documents

Pro gradu –tutkielma, 81 sivua  
Huhtikuu 2012

---

Tässä pro gradu –tutkielmassa tarkastellaan käyttöohjeiden visualisointia. Tutkimusidea nousi tekniseen dokumentointiin erikoistuneen firman, Lionbridgen, tarpeista. Tutkimuksen tarkoituksena on selvittää, miten tekstiä ja kuvia olisi hyödyllistä yhdistää teknisissä dokumenteissa ja miten teknisiä dokumentoijia voidaan auttaa tunnistamaan, milloin ja millaisten kuvien käyttö on käyttäjien kannalta kaikkein hyödyllisintä.

Tutkimuksen teoreettinen viitekehys koostuu kahdesta pääosasta: ensimmäisessä osassa käsitellään teknisen dokumentoinnin alan syntyä ja kehitystä sekä multimodaalisuutta ja sen yhteyttä tekniseen dokumentointiin. Toinen osa puolestaan keskittyy tekstin ja kuvan yhdistämisen teoriaan: siinä keskitytään teknisten dokumenttien visualisointiin vaikuttaviin tekijöihin, tekstin ja kuvan erityispiirteisiin sekä tekstin ja kuvan välisiin suhteisiin.

Tutkimuksen aineistona käytetään Lionbridgen teknistä dokumentaatiota. Työssä analysoidaan kahta käyttöohjetta: Nokian *Lumia 800* älypuhelimien käyttöohjetta ja Lionbridgen sisäisen projektinhallinnan hyötyohjelman, *Geminin*, käyttöohjetta. Nämä valittiin sen perusteella, että ne edustavat eri tuotetyyppejä: *Lumia 800* on konkreettinen esine ja *Gemini* on abstrakti tietokoneohjelma. Analyysissa tarkastellaan ensin ohjeissa esiintyvien kuvien käyttötarkoituksia ja sitten näiden kuvien yhteyttä tekstiin. Työssä perehdytään myös näiden kahden osion välillä oleviin yhteyksiin, eli siihen, yhdistyvätkö tietyntyyppiset kuvat tekstin kanssa aina samalla tavalla.

Tutkimus osoittaa, että kuvien käyttötarkoitukset ja yhdistämistavat tekstin kanssa poikkeavat käyttöohjeiden välillä. Erot dokumenttien välillä johtuvat pääosin siitä, että käyttöohjeet kuvaavat kahta täysin erilaista tuotetyppiä, älypuhelimia ja tietokoneohjelmaa. Vertailemalla tutkittavia käyttöohjeita pystyttiin selvittämään, miten tekstiä ja kuvia on hyödyllistä yhdistää. Analyysissa kävi ilmi, että teksti toimii hyvin kuvan tukena *Lumia 800* älypuhelimien käyttöohjeessa, kun halutaan vahvistaa tekstillä kuvattua toimintaa. *Geminin* käyttöohjeessa kuvat puolestaan tukevat hyvin asioita, jotka ovat luonteeltaan hyvin abstrakteja. Kuvattaessa avaruudellisia suhteita tai sijaintia teksti ja kuva toimivat molemmissa käyttöohjeissa parhaiten toisiaan täydentävinä. *Geminin* käyttöohjeessa toisiaan täydentävä teksti ja kuva toimivat hyvin myös tilanteissa, joissa halutaan vähentää käyttäjän työmuistin kuormitusta. *Lumia 800* älypuhelimien käyttöohjeessa kuvia käytettiin myös uusien kappaleiden alussa johdattelemaan lukijaa uuteen teemaan. Näiden kuvien todettiin myös auttavan käyttäjää navigoimaan käyttöohjeen sisällä. Mielenkiintoinen aihe jatkotutkimuksia ajatellen olisikin selvittää, miten käyttäjät reagoivat näihin erilaisiin teksti–kuva-yhdistelmiin, joita tutkimus tuo esille.

Avainsanat: visualisointi, käyttöohje, multimodaalisuus

# Table of Contents

<b>1 Introduction</b> .....	<b>1</b>
1.1 The Aim of the Study .....	2
1.2 The Structure of the Study .....	5
<b>2 Theoretical Background</b> .....	<b>6</b>
2.1 The Field of Document Design .....	6
2.2 Multimodality.....	8
2.2.1 What is Multimodality? .....	9
2.2.2 The Effects of Multimodality on Document Design.....	12
<b>3 Interplay of Text and Images</b> .....	<b>14</b>
3.1 Visualisation of Documents .....	14
3.2 Characteristics of Text and Images .....	20
3.3 Integration of Text and Images .....	28
<b>4 Material and Methods</b> .....	<b>35</b>
4.1 Lumia 800 and Gemini .....	35
4.2 Method .....	37
4.2.1 Images .....	37
4.2.2 Relationships between Text and Images .....	38
<b>5 Analysis</b> .....	<b>39</b>
5.1 Use of Images in the Manuals .....	39
5.1.1 Lumia 800 .....	40
5.1.2 Gemini.....	48
5.2 Integration of Text and Images in the Manuals .....	57
5.2.1 Lumia 800 .....	57
5.2.2 Gemini.....	65
<b>6 Conclusions</b> .....	<b>75</b>
<b>Works Cited</b> .....	<b>78</b>

## List of Figures

Figure 1: Factors that affect visualisation of documents .....	15
Figure 2: Functions of images in <i>Lumia 800</i> 's user manual .....	40
Figure 3: Functions of images in <i>Gemini</i> 's user manual .....	51
Figure 4: The relationships between text and images in <i>Lumia 800</i> 's user manual .....	57
Figure 5: The relationships between text and images in <i>Gemini</i> 's user manual.....	65
Figure 6: Guidelines for the integration of text and images in hardware documentation.....	73
Figure 7: Guidelines for the integration of text and images in software documentation.....	74

## List of Tables

Table 1: The characteristics of text and images .....	27
---	----

## List of Examples

Example 1: Image that expresses spatial relationships 1 ( <i>Lumia 800</i> , page 6).....	41
Example 2: Image that expresses spatial relationships 2 ( <i>Lumia 800</i> , page 13).....	42
Example 3: Image that reinforces verbal description ( <i>Lumia 800</i> , page 20) .....	43
Example 4: Instructions for swiping and zooming ( <i>Lumia 800</i> , page 21).....	44
Example 5: Orienting image ( <i>Lumia 800</i> , page 31).....	45
Example 6: Image that verifies screen state ( <i>Gemini</i> ) .....	51
Example 7: Locating screen capture 1 ( <i>Gemini</i> ).....	53
Example 8: Locating screen capture 2 ( <i>Gemini</i> ).....	54
Example 9: Screen capture that helps to build a mental model 1 ( <i>Gemini</i> ).....	55
Example 10: Screen capture that helps to build a mental model 2 ( <i>Gemini</i> ).....	55
Example 11: Supplementary relationship 1 ( <i>Lumia 800</i> , page 20).....	58
Example 12: Supplementary relationship 2 ( <i>Lumia 800</i> , page 46; emphasis mine).....	59
Example 13: Complementary relationship 1 ( <i>Lumia 800</i> , page 8) .....	60
Example 14: Complementary relationship 2 ( <i>Lumia 800</i> , page 12) .....	61
Example 15: Stage-setting image 1 ( <i>Lumia 800</i> , page 15) .....	62
Example 16: Stage-setting image 2 ( <i>Lumia 800</i> , page 18) .....	62
Example 17: Redundant relationship ( <i>Lumia 800</i> , page 10).....	64
Example 18: Supplementary relationship ( <i>Gemini</i> ).....	66
Example 19: Insufficient use of the supplementary relationship ( <i>Gemini</i> ) .....	67
Example 20: Efficient utilising of the complementary relationship ( <i>Gemini</i> ).....	68
Example 21: Complementary relationship ( <i>Gemini</i> ) .....	69

# 1 Introduction

“I never read, I just look at the pictures.”  
-Andy Warhol

“A picture is worth a thousand words” is an age-old adage that is widely known in different cultures. But despite the fact that the adage is old, we could claim that only in today’s visual world we can really understand the significance of it. We encounter images everywhere – at home, in the workplace, in the streets, and in the virtual world. Thus, images are an important part of our everyday communication.

This growing use of visuals in communication has affected the field that this study focuses on. This field is called *document design* and it is in Karen Schriver’s (1997, 10) words “concerned with creating texts (broadly defined) that integrate words and pictures in ways that help people to achieve their specific goals for using texts at home, school or work.” Although Schriver (1997, 20) mentions both words and images in her definition, document design has for a long time focused on writing. However, the growing importance of visuals in our culture also challenges document designers to inspect the relationships between text and images more closely. Mitch Klink (2000), a veteran of document design, states that until recently he has primarily thought of himself as a writer, and that he has overlooked the impact of visual media on our culture. Klink has realised that he must be willing to re-invent himself to meet the evolving needs of a visually oriented society.

Klink (2000) points out that today’s developments in communication technology have given us plenty of new possibilities for the exchange of information. However, these new opportunities have also introduced new challenges to document designers. As our world becomes more and more complex, a creative use of images will be a key factor in effective communication. Thus, document designers are not supposed to concentrate merely on writing: they need to be able to decide what kind of combinations of text and images they will use in different contexts. As Jeffrey Donnel (2005, 269) aptly points out, the challenge of document design is not simply to choose a suitable

verbal genre for particular documents, but to simultaneously create and integrate coherent visual and verbal discourses.

Claire Harrison (2003, 47) notes that in this new multimodal communication environment communicators face three significant challenges:

1. To understand how text and still images work together to make meaning together for readers/users.
2. To know when still images enhance or detract from text, and vice versa.
3. To be able to effectively discuss the issues of multimodal communications with other members of the document's production team.

These challenges validate the need for this study: a systematic study on the interplay of text and images is needed so that document designers are ready to meet these new challenges, and especially that they are ready to develop themselves as professional communicators. With the help of this study, I want to find out how document designers can take advantage of images in documents: what are the strengths of text and images and how can these strengths guide document designers to make decisions about the integration of text and images?

## **1.1 The Aim of the Study**

The idea for this master's thesis came from Lionbridge Technologies, Inc. (hereafter Lionbridge). Lionbridge is a company that provides translation, localisation, internationalisation, interpretation, content development, software development, and software testing services. When I started to think about the topic for my thesis I tried to find a company that would share my interest in *multimodality* (see chapter 2.2), and Lionbridge's offer sounded interesting and timely. They proposed that I could study the visualisation of technical documentation. We discussed the topic together and formulated one main question on which this study will focus: How to integrate text and images in technical documentation? In other words, how can document designers recognise when and what kinds of

images are the most useful for the users of the documentation, bearing in mind the operative function of documentation.

As the focus will be on the interplay of text and images, the actual production of images and image tools will be left out of the study. Schriver (1997, 364) points out that professionals are interested in knowing more about the interplay of text and images but the literature available has been disappointing: the literature on the interplay of text and images tends to focus on how technology can be used to combine these two instead of on how to choose the type of words and images that the user needs and wants. The same lack of literature on the integration of text and images seems to be a problem for document designers even 10 years after Schriver's comment. Eva Brumberger (2007, 376) remarks that in the field of document design, scholarly conversation has focused on the practice, research, and pedagogy of visual rhetoric. However, in spite of the conversations on visual rhetoric, visual thinking has received relatively little attention within the field. Brumberger (2007, 376) states that "if our programs produce students who can think verbally but not visually, they risk producing writers who are visual technicians but are unable to move fluidly between and within modes of communication." Thus, there is an urgent need for studies that will focus on the integration of text and images in technical documentation.

The fact that Lionbridge, which is a global provider of language services, wants more information specifically on visualisation is indicative of the fact that there is a need for this type of study in the language industry in general. Actually, one by-product of my thesis will be teaching material on visualisation for other document designers. When my thesis is ready I am going to hold two seminars on visualisation of documents for the employees of Lionbridge. I hope that my seminars will encourage document designers to discuss the issues of multimodal communications so that they can develop their skills in using images in technical documents.

The material analysed in this thesis will be technical documentation provided by Lionbridge. I will analyse two user manuals: the one is the user manual of a smart phone, *Lumia 800*, and the other is the user manual of Lionbridge's internal project management system, *Gemini*. These

specific user manuals were chosen, firstly, because they represent different product types. *Lumia 800* is a concrete object, a phone, and its user manual includes line drawings of the equipment. *Gemini*, on the contrary, is software and thus it is clear that the user manual includes screen captures. Consequently, I assume that the use of images in the user manual of *Lumia 800* will differ from the images used in *Gemini*. As chapter 3.1 will show, software and hardware documentation often use images for different purposes. Software documentation focuses on helping users to understand abstract concepts, whereas hardware documentation also offers information on the concrete product. Secondly, these user manuals were chosen because they do not contain confidential information so it is easier to give examples of the material throughout my study.

The method to be used in this study is two-fold. I will begin by analysing the use of visuals in the two manuals. My intention is to find out for what kinds of purposes images are used, in other words, what their function in the manual is. After analysing the visuals, I will move on to the relationships between text and images. I am going to conduct a contrastive analysis of text and images with the help of a model that Schriver (1997) introduces in her book *Dynamics in Document Design*. Schriver's (1997, 412–413) model includes five ways in which text and images can be integrated: *redundant*, *complementary*, *supplementary*, *juxtapositional* and *stage-setting*. Each of these relationships represents different ways in which images and text can interact with each other. To give an example, an image can give the same information as the accompanying text (redundant relationship) or the image may be the dominant mode while text only elaborates it by providing additional information (supplementary). I will discuss the relationships between text and images more thoroughly in chapter 3.3.

After the analysis of both sections my aim is to see whether there is a connection between the functions of images and the relationships they form with the text. That is to say, whether some specific types of images always form the same kind of relationship with the text, for example.

## **1.2 The Structure of the Study**

I will begin this study by introducing the theoretical background in Chapter 2: I will give a brief introduction to the field of document design as well as discuss multimodality and its effects on document design. In the third chapter, I will discuss the general factors that affect visualisation of documents, the characteristic features of text and images and the integration of text and images.

After these theoretical chapters, I will move on to present the material and the method used in this study in Chapter 4. In Chapter 5, I will report on the results of the analysis of the two manuals. I will begin by reporting on the functions of images in both of the documents, and then, I will focus on the integration of text and images. In the final chapter, I will draw some conclusions of the results and evaluate the successfulness of the study. My aim is also to propose some recommendations for further studies that concern the visualisation of documents.

## 2 Theoretical Background

The idea of this chapter is to provide a framework which functions as a base for this whole study. I will begin by introducing the field of document design. My intention is to describe the factors that have shaped the field of document design and highlight the factors that are relevant for this study. After describing the field, I will focus on multimodality. I will give a general description of multimodality and discuss the effects of multimodality on document design.

### 2.1 The Field of Document Design

In this chapter the field of document design will be described more closely. In my opinion, without a general picture of the field, a detailed analysis on the interplay of text and images is not possible. Thus, the idea of this chapter is to provide background information on document design, the focus being on the use of text and images.

As noted in Introduction, Schriver (1997, 10) points out that the aim of document design is to produce texts that help people to achieve their goals. With the help of technical documentation, people learn, use technology, and get their work done. A document can be a user instruction for a mobile phone or maintenance manual for a jet engine, for example. But whatever the type of the text is, according to Schriver (1997, xxiii), documents are created to be useful. That is to say that good documents get us to read them and they communicate with us. Schriver (1997, xxiii) points out that the purpose of document design is to explore how good writing and visual design can improve the documents with which people deal.

The need for document design rose from social and technological forces. In the 20<sup>th</sup> century, document design developed most dramatically in industrialised, market-oriented countries. Many types of documents were needed to help people to complete their day-to-day activities in the workplace or in the home. Of course, the nations differ in the conditions that led to the development of document design, but some of the main forces unquestionably included growing consumerism

and new technological inventions. People needed documents so that they could learn how to use the new products they were now able to buy. (Schrivver 1997, 16.)

However, some remarkable changes have also happened in the field of document design in the last few decades. Once again, new technologies have given us plenty of opportunities to make documents even more effective. According to Schrivver (1997, 362), originally those who wrote documents were viewed primarily as “word people”. Document designers’ task was to write documents that are useful to users. They needed to be fluent in written communication so that they could make sure that the user understands the message. However, the situation has changed. According to Tiffany Portewig (2008, 333), since the 1980s, scholars have highlighted the importance of visual communication in document design. Being a proficient writer is no longer enough for professional communicators, but they need to be effective visual designers as well.

Thus, we have moved from a world of print media to the world of electronic media. According to Richard Johnson-Sheehan (2002, 79), the invention of the personal computer can be seen as “the catalyst for finally shifting the literal culture into a visual one, much as the printing press was the catalyst for shifting the oral culture to a literal one.” Johnson-Sheehan (2002, 75) argues that if the medium is electronic, then the primary rhetorical element is visuality. This, of course, creates challenges for document designers, who are used to working with words.

This growth in importance of visual communication has even caused difficulties in the naming of the field. The fact that I have chosen to use the term document design is not as simple as it perhaps seems to be, and that is why it is important to validate the choice. Schrivver (1997, 4) remarks that there is no perfect name for the field called document design and that is why it is often misunderstood. The naming of the field is important because it provides a common language and helps to describe the territory of the field. It also gives the members of the field a sense of identity and at the same time it gives outsiders an idea of what the field is about. In my opinion, the term that people prefer to use tells about their attitudes towards the field and especially what they think the purpose of the field is.

To begin with, Schriver (1997, 4) states that the term document design has been criticised because of the connotations of those two words, “document” and “design”. According to Schriver, the word “document” strikes a negative chord for many people, because they are accustomed to associating the word with hard-to-understand tax forms and cryptic instruction guides, for instance. It is also disputable whether the word “document” is sufficient in today’s world of multimedia where documents have a clearly different look than several decades ago. On the other hand, “design” is not unproblematic either. Many people would connect the term with products of architects, product designers, or fashion moguls rather than technical documents.

Schriver (1997, 6) argues that these difficulties in defining the words “document” and “design” have made some writers and designers to think whether some other name would be more appropriate. There are several different kinds of names for the field which all emphasise different characteristics of the field, for instance, *information design*, *communications design*, *professional communication*, and *technical communication*.

Schriver (1997, 10) herself has chosen to use the term document design because, in her opinion, “it suggests the *act* of writing and designing – the process of bringing together words and pictures.” I have decided to use the term as well because I think it suits my purposes better than well. In my opinion, the name of the field should emphasise that in today’s visual world document designers are not just creating text but also designing images that function with the text.

## **2.2 Multimodality**

The aim of this chapter is to provide such information on multimodality that will support my analysis of the documents. Firstly, I will define the term multimodality and introduce some of its major features. The focus will be on multimodal texts in the electronic era, because both of the studied user manuals are electronic documents. Secondly, I will discuss the significance of multimodality in the field of document design: how does the growing use of multimodal

communication change the field of document design and how should document designers react to these changes?

### 2.2.1 What is Multimodality?

An important part of the theoretical background of my study falls under the term multimodality.

Gunther Kress & Theo Van Leeuwen (2001, 20) define multimodality as: “the use of several semiotic modes in the design of a semiotic product or event, together with the particular way in which these modes are combined.” Those modes can, for instance, reinforce each other by saying the same thing in different ways or complement each other. Sometimes one mode is the dominant one and the other supports it, or they may be equally important. However, the main principle of multimodality is that different modes make meaning together.

Thus, multimodality refers to the mixing of different modes. There are many different types of modes, but this study will concentrate on two modes, namely images and text. According to Jeff Bezemer & Gunther Kress (2008, 171), “[a] *mode* is socially and culturally shaped resource for making meaning.” Bezemer and Kress (2008, 171) list, for example, image, writing, layout and moving images as different modes. People create meanings by combining these modes that all have differing *modal resources*. Writing, for instance, has syntactic, grammatical, graphic (such as font type) and lexical resources, whereas resources of images include spatial relation and position of elements in a framed space, size, colour, and shape, for instance. Because of these differences between modes they can be used for different kinds of semiotic work. That is to say, different modes have different potentials and constraints in making meaning. Consequently, when document designers make decisions about the integration of text and images, they need to be aware of the potentials and constraints that each type of mode has. I will return to the modal resources of text and images in chapter 3.2.

Bezemer and Kress (2008, 172) point out that another important term that has to be considered together with mode and modal uses is the *medium*. A medium always has a material and

social aspect. Bezemer and Kress (2008, 172) remark that “[m]aterially, medium is the substance in and through which meaning is instantiated/realized and through which meaning becomes available to others. . . .” According to the definition, print, book, screen and “speaker-as-body-and-voice” are all different kinds of material media. On the other hand, socially, a medium can be considered to be the result of semiotic, sociocultural, and technological practices, such as film, newspaper, billboard, radio and television. Consequently, the joint effect of mode and medium makes multimodality possible. Carmen Maier, Constance Kampf and Peter Kastberg (2007, 456) concisely point out that “a medium can contain multiple modes of communication, and thus be multimodal.”

The opposite of multimodal is *monomodal*, which simply means that only one mode, method, system etc., is used. In western cultures, textual monomodality has been for a long time considered to be somehow better than textual multimodality. The most valued and important forms of text have been those that do not have any images but just text – for example novels and scientific reports. However, the situation is not similar today. The dominance of monomodality has begun to crack and the use of multimodal texts has increased significantly in recent years. (Lehtonen 2002, 46–47.)

However, according to Kress & Van Leeuwen (2001, 1), it is not only the mass media, magazines and comic strips that break the dominance of monomodality but also documents produced by corporations, universities, and government departments. Eija Ventola, Charles Cassily and Martin Kaltenbacher (2004, 1) remark that the emergence of new media has forced scholars to think about the characteristics of different modes and the way those modes function together semantically and ways in which they can be combined.

Mikko Lehtonen (2002, 47) states that although the multimodality of our culture has become more important and more visible in the past few years, multimodality is by no means a new phenomenon. As long as there have been human cultures, there has been multimodality, too. Even when we are talking with each other, we rarely rely purely on verbal means, but we often use several non-verbal gestures and forms of body language. As a matter of fact, Ventola et al. (2004, 10) argue that purely monomodal text has always been an exception and the core practice in

communication has been primarily multimodal. Kress (2000, 187) takes the exceptionality of monomodality even further by arguing that all texts are multimodal. Kress states that no text can exist in a single mode although one modality can be the dominant one. The idea behind Kress' claim is the fact that even in those texts that do not include any images there are visual elements such as font and spatial arrangement that make the texts multimodal. Consequently, this is simply a matter of how thoroughly the term is defined. In this study, however, such an extensive definition of multimodality is not used.

So if multimodal texts have always existed, what makes them so important today? Ventola et al. (2004, 1) state that despite the fact that multimodality has always been present in most of the communicative contexts in which humans engage, it has for a long time been ignored. However, the developments in technology have made it easier to combine different modes, and that forces scholars to think about the particular characteristics of these modes and the way in which they function semantically in the modern discourse worlds. Kress (2003, 5) states that with print-based technology, the production of text was made easy but the production of images was more difficult and that is why images were not used so often. However, in today's technologically developed world, multimodality is made easy, usual and natural by the new technologies we have in hand.

These technological changes have changed our communication environment. According to Kress (2003, 35), it is no longer possible to treat literacy as the main means for communication. Other modes are there as well, and in many environments they can be even more prominent and significant ways of communication than written words. Lehtonen (2002, 56–59) also states that we should not think that printed text will always be the most dominant form of media in teaching and in research. Conversely, we should be ready to deal with the visualisation of our culture. According to Lehtonen (2002, 46–59), the economic and technological changes in the world may make the visual and multimodal texts dominant. It is more and more common that images are not just decorations inside the text paragraph, but the image actually becomes the most important thing and the text serves only as a commentary to the image.

However, Ventola et al. (2004, 10) point out that although the achievements in the research on multimodality have been quite remarkable, studies on the interrelations between various modes is underrepresented. It seems that we know more about the function of individual modes than about how they interact together and how they are organised in text and discourse. In my opinion, that is why more research on multimodality is clearly needed in order to help people to take advantage of the positive effects of multimodality.

### **2.2.2 The Effects of Multimodality on Document Design**

The extensive use of multimodal communication in today's media has also brought challenges to document designers. Klink (2000) remarks that changes in the way our culture chooses to share information affects the role that document designers play in the process: to be able to survive as a document designer in the world of multimedia requires adaptability and a wide knowledge base. Kress and van Leeuwen (2001, 47) points out that digital technology has now made it possible for one person to manage several modes and make multimodal products single-handedly. Basically this means that document designers can combine words, images, video, and audio as they wish. William Gribbons and Arthur Elser (1998, 467) make an apt comparison by saying that “[j]ust as technical communicators rose a decade ago to the challenges of typography, illustration, and page layout; communicators of the 21st century must meet the challenges of visualizing information.” This shows that document designers need fluency in both visual and verbal thinking and consequently they need to be able to create documents that include both verbal and visual information. However, today the problem is not how to make videos, audio files and images but how to make these modes work effectively together.

Kress (1998, 67) raises an important question concerning the growing amount of visual information in information technology. Kress asks whether images and language merely co-exist or do these two semiotic modes interact with each other. And most importantly if those two modes actually interact with each other, what are the consequences? According to Kress (1998, 72–73),

images and language are not just coexisting but, it seems that there actually is a strong interaction between those two modes. Consequently, if language is no more the central semiotic mode, then theories of language are not alone sufficient to explain the communicational landscape. Thus, a theory is needed that deals adequately with the integration of different modes in multimodal texts.

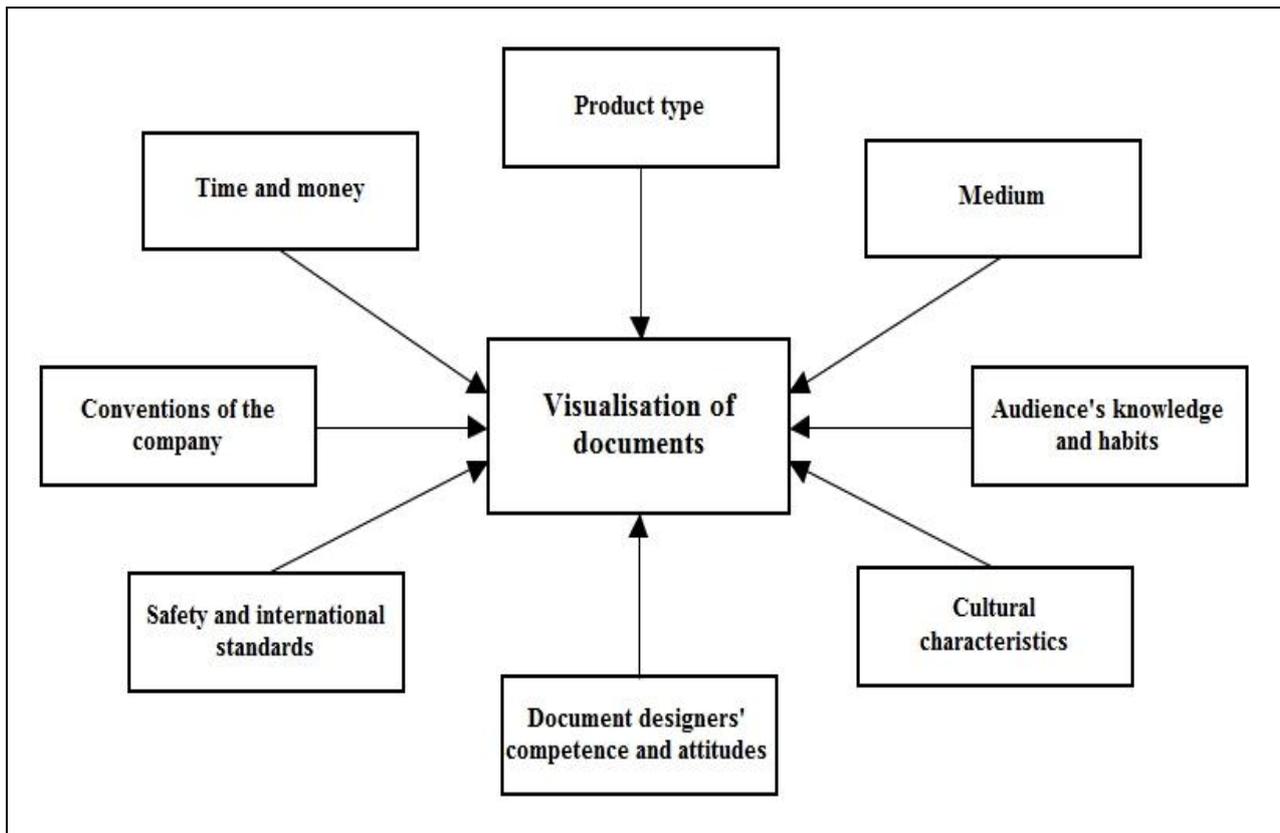
The point that Kress brings forward should be taken under careful consideration in the field of document design. According to Maier et al. (2007, 453), a multimodal analysis can help document designers to understand the needs of future generations that have grown up in front of the computer screen. That is to say that document designers need to be able to respond to the needs of people who are becoming more and more multimodally literate. Maier et al. (2007, 454) point out that *multimodal literacy* is the key concept in post-modern audience analysis. Consequently, it is crucial that document designers, who should always keep the audience in mind, understand the impact of growing multimodal literacy among people. According to Maier et al. (2007, 457), “[m]ultimodal literacy means that the audience is savvy enough to not only understand pictures and words, but also understand combined meaning that is shared and/or multiplied across modes of communication in any given multimedia publication.” In chapter 3.3 I will concentrate on introducing some models that have been used to examine multimodality.

## **3 Interplay of Text and Images**

Now that I have discussed multimodality in general, I will move on to a more specific type of multimodality: the interplay of text and images. However, before moving on to describing the use of text and images, I will introduce the factors that affect the visualisation of documents. After that, I am going to concentrate on the differences between text and images. Finally, I will focus on the relationships between text and images and present Schriver's (1997) model of how text and images can be integrated.

### **3.1 Visualisation of Documents**

As we just discussed, the general view seems to be that adding visuals into documents is a crucial part of making documents effective. However, it is important to take into account all the prerequisites that influence the choices that document designers need to make when they visualise documents. In my opinion, it is necessary to identify the external factors that affect visualisation, before moving on to a more detailed description of text and images. The Figure 1 compresses all those factors, which will be discussed in this chapter:



**Figure 1: Factors that affect visualisation of documents**

As Figure 1 shows, many factors need to be taken into account when visualising documents. The factors can be roughly divided into four groups: factors that have to do with the company, the users, the document itself and the document designers. First of all, Portewig (2008, 430) points out that resources such as **time and money** play an important role in the use of visuals in documentation. Adding images to documents is not cheap (especially if you need to hire someone outside your own company to do it) and it is also time-consuming. It would be ideal if document designers were able to use images as they wished, but the reality is that companies have budgets and timeframes, which determine how much the documentation can cost and how much time document designers can use on it. Companies may also have different kinds of **conventions** that document designers need to take into account when they visualise information. Portewig (2008) has studied the role of invention for visuals in the workplace by interviewing document designers in three different companies. One factor that document designers mentioned was conventions. According to

interviewed document designers, there are often such components in some of the products that are always illustrated. In addition, Portewig's (2008, 339) study shows that **safety and international standards** also influence the choices that document designers make when they visualise documents.

Considering when it is appropriate to visualise documents, document designers also need to keep in mind the ultimate purpose of documentation. The purpose of documentation obviously is that the users complete the task they want to complete. Thus, one factor that affects the visualisation process is **audience's knowledge and habits**. According to Schriver (1997, xxiii), users deserve documents that meet their needs, and it is document designers who play a central role in making this happen. Consequently, in addition to the company's needs, document designers need to keep the audience in mind while they make decisions about the content of the document. The fact is that document designers always have to balance between the users' needs and the company's needs.

Schriver (1997, 166) states that document designers must use visual and verbal language that connects with the users' knowledge, experience, beliefs, and values. Most probably, document designers choose different kinds of combinations for experts and for novices. Inexperienced users may need more supplementary images to help them to understand the text than experts, for example. Of course, the most appropriate format for presenting information also depends on the complexity of the task. However, what kinds of things can be categorised to be complex depends on the users' knowledge.

So, according to Schriver (1998, 365–367), when users interpret visual and verbal language, their unique experiences affect the process. In order to understand what happens during the interpretation of documents, it is important to consider how people read. Reading is a complex knowledge-driven and text-driven process. The users' interpretation of the text depends on the evidence they get from multiple interacting cues, text-driven and knowledge-driven. Knowledge-driven cues refer the things that the user brings to bear during interpretation: knowledge, experience, feelings, social awareness, and culture, whereas text-driven cues refer to the users' interaction with visual or verbal signs. These cues include, for example, word meanings, sentence

structures, images, charts, and so on. Consequently, as Schriver (1997, 368) points out, document designers need to make textual moves “that will help users with both their knowledge-driven and text-driven constructions of the text and graphics.”

Moreover, Schriver (1997, 164–165) states that one general principle about audience that has to be taken into account in document design is the fact that people prefer not to read unless they have to. Skilled users have strategies which help them decide what to browse, skim through, examine carefully, or skip altogether. That is why it is important that document designers structure the document so that the main ideas catch the attention of busy users. Arguably, images are good at drawing users’ attention. William Pfeiffer (2000, 399) points out that images, font styles and colour are “grabbers”: they engage users’ interest. To use Pfeiffer’s example, if you have three reports on your desk and you must quickly choose which one you will read first, you will most probably choose the one that has the most distinctive look. In addition to attractiveness, images often create a feeling that the information is important. Consider that you are reading manual that mainly consists of textual information. If you suddenly see an image in the manual, you will presumably think that the information it presents is somehow important. According to Lu et al. (2009), document designers frequently use images to present important information. On the other hand, also end-users tend to search for images and figures in documents. Consequently, I would argue that images have an essential function when document designers want to guide the users to inspect the most important parts of the document. When the fact is that users of documents read only as much as they have to, it is important that document designers clearly indicate what they think is the most relevant information, and consequently, what they want the user to inspect with careful attention.

As we can see, every user’s personal traits affect how they read and understand documents. However, a specific group of people also share some **cultural characteristics** that should guide the choices that document designers make. Schriver (1997, 364) emphasises the importance of understanding the following paradox: “Reading is a social act in that it depends on a community that shares meanings; yet it is also an individual act in that it depends critically on the reader’s

unique knowledge, attitude, and values.” That is to say that document designers should be able to take both individual differences and cultural similarities into account when creating technical documentation.

Harrison (2003, 48–49) states that because all communities are unique, the signs that are used in one community may not be used in another. She points out that the colour red is a sign of mourning for people in Ivory Coast, whereas in India it symbolises procreation and life. William Horton (1992, 193) also gives an example of a symbolic gesture that is understood differently in different cultures. Horton states that “the thumbs-up gesture” that is used to hitch a ride or signal that everything is OK in the United States (and in Finland as well) is considered to be an obscene gesture in many Mediterranean countries.

In addition to these kinds of symbolic differences between different cultures, the number of images and the type of images in user manuals also tend to vary. Wang Qiuye (2000), for example, has studied the differences in the use of images between Chinese and American scientific and technical communication. Qiuye (2000, 554) state that it may be difficult to a user from one culture to approach the visual language of another. The aim of Qiuye’s study was to find out how one culture can differ from another culture in the use of visual communication.

The results of the study showed that there actually are some cultural differences in the use of visuals between these two countries: the images in American manuals emphasise task performance and they are larger in size and more detailed than the images in Chinese manuals. On the other hand, in Chinese manuals, however, most of the images are used when introducing the product information and there are not so many images accompanying the steps that help the users to complete their tasks. Chinese manuals also tend to provide more contextual information in the form of images, while American manuals tend to be more direct. So if document designers need to write for international audiences, they need to be aware of the fact that visual information can also have different meanings in different social and cultural contexts.

So far I have discussed factors that fall under two main categories: company and audience. But there are also factors that relate more closely to the actual document that is being created: the **type of product** that the documentation addresses and the **medium** by which the information is communicated. To begin with, different types of images are used in the documentation of different types of products. As Elaine Lewis (1988, 239) states the type of image to be used depends on the characteristics of the object that is being visualised. To give an example, software documentation often includes a number of screen captures, whereas in hardware documentation, photographs or line drawings of the product are the more natural ones. However, software and hardware documentation also differ with regard to the purposes for which images are used. According to Lewis (1988, 245), images in hardware documentation are especially useful in representing equipment, systems, and components. Lewis points out that images can reinforce the verbal descriptions of the hardware and enhance comprehension of assembly and maintenance tasks. Users remember descriptions with images better than text alone versions. Lewis (1988, 242–243) states that describing conceptual processes and procedures is an important function of the images in software documentation because images clarify abstract content. They enhance understanding and help to remember the information.

The choice of media also has an effect on the visualisation of documents. Using images in online documentation is cheaper compared with printed versions, because images do not have to be printed but just displayed. More importantly, the structure of online documentation is fundamentally different from that of print documentation, which of course affects the visualisation process. Pfeiffer (2000, 596) remarks that online documents allow the user to interact with the document in a way that could never be done with paper. Users can often use search engines or online indices in order to find information they need. In addition, they can use hyperlinks to navigate between different topics in documents. An important feature of online documentation is its use of multimedia: online documents can easily include sound, video, animation, and images.

Finally, it can be noted that the professional **competence** and especially **attitudes of document designers** have a considerable effect on the visualisation process. The fact that many designers see themselves as writers who produce text rather than designers who integrate different modes of communication may downplay the creative use of visuals. Thomas Williams and Deborah Harkus (1998, 33) remark that document designers are generally reluctant to use images and prefer using words instead. Williams and Harkus think that this behaviour arises partially out of habit and partially out of a belief that words are the most appropriate format to convey serious discourse. Of course, the situation can be slightly different today: document designers may have a more positive attitude towards image integration because the use of images in documents have become more and more common. However, as I stated earlier, I believe that because Lionbridge wants more information specifically on the integration of images in documents, it can be concluded that there still is some uncertainty in the use of visuals among document designers.

Harrison (2003, 46), who herself is a document designer, states that those who create documents are trained and practiced in the use of words. She remarks that when she needed to decide which image(s) would be best for some specific purpose, she generally relied on her “gut feeling”, which made her feel rather uncomfortable. I believe that many document designers in the field have found themselves in the same kind of situation. The purpose of this thesis is to make such situations easier to cope with by providing basic instructions for the effective integration of words and images.

### **3.2 Characteristics of Text and Images**

Throughout this study I have discussed the importance of visuals in today’s documents. However, it is important to keep in mind that although “one picture can tell us more than a thousand words”, it does not mean that by increasing the amount of images we can automatically make the documents better. Charles Kostelnick (1994, 91) remarks that technology has never in the history of business and document design given us such powerful design tools and left us so ill prepared to use these

tools intelligently. He says that “although we now largely recognize this new visual landscape, we have little perspective with which to explore or to understand this new territory or to exercise the freedom it affords us to compose documents visually.”

The problem for today’s document designers is that they do not have guidelines for choosing appropriate images to accompany their texts. Russell Willerton (2005, 3) states that it is not easy for document designers to incorporate more visuals into communication, because they lack guidelines for selecting and composing effective images. So are there any simple rules that document designers can use when they make decisions about how to combine text and images? What kind of information could best be conveyed via images and respectively via words, bearing in mind all the factors presented in the previous chapters?

Presumably, without simple rules of thumb that document designers can use as a support for their choices, document designers presumably concentrate merely on the mode with which they are most familiar: writing. In the study carried out by Portewig (2008), document designers were asked how they decide what information should be communicated visually in technical documents. Portewig (2008, 338) states that she repeatedly got the same kind of comment from the document designers she interviewed: they do not think so much about the decisions they make when they use visuals. The document designers used visuals when they had difficulties explaining something with words. This comment neatly compresses the dilemmas of using images in document design. One dilemma seems to be the fact that images in documents are often treated as subordinate to the text. Jeffrey Donnel (2005, 241), for example, points out that the textual approach to document design is presumably based on an untested assumption that text functions as a primary means of communication, while images’ function is to support the text. In addition, the comment proves that guidelines for selecting visual and verbal content are needed to help the document designers to cope with the growing demand for multimodal communication.

According to Ronald Fortune (2002, 103), in order to understand how words and images interact in an electronic document we need to recognise how they differ fundamentally. Fortune

(2002, 105) claims that problems will undoubtedly arise when those who create documents do not understand how words and images work alone and together. That is why I think it is reasonable to begin by comparing these two modes of communication before moving on to the interaction between them.

Williams and Harkus (1998) provide some practical guidelines for making choices between visual and verbal communication by comparing and contrasting images and words. Williams and Harkus (1998, 33–34) quote Gavriel Salomon (1979)<sup>1</sup> by saying that images and text are both symbol systems and that different symbol systems can best represent different kinds of ideas. They give a concrete example of this idea by asking readers to recall a situation when they got frustrated when reading a complicated verbal description. In these kinds of situations people often desperately want images instead of a cumbersome text. Conversely, some ideas are more easily communicated with text: anyone who has played *Pictionary*, a game where you have to explain things by using visual language exclusively, will admit that there actually are some ideas that are very difficult to communicate with images.

However, according to Williams and Harkus (1998, 34), despite the fact that some ideas are highly challenging to represent in some symbol systems, the ideas that these symbol systems can represent overlap considerably. This means that communicators must choose the best way to deliver the information among the available modes of communication, usually the best way being the one that is to most useful given the users' needs and preferences. The challenge for communicators is that the "correct" choice is not always so explicit and the consequences of a "poor" choice can make the users' task more difficult. In their article Williams and Harkus discuss some of the most fundamental differences between text and visuals and the effects those differences have on the choices that communicators need to make. Those differences include:

---

<sup>1</sup> Salomon, Gavriel. 1979. *Interaction of Media, Cognition and Learning*. San Francisco, CA: Josey-Bass.

1. differences in how symbols in each system evoke their referents;
2. differences in the nature of the referents they evoke;
3. differences in the structure each symbol systems imposes on the information it carries; and
4. differences in the degree to which information carried in either system can be processed perceptually.

Firstly, according to Williams and Harkus (1998, 34–36) images and text are different in the way in which they evoke their referents. Images and words are both “coding elements” that substitute other things, their referents. The relationship between words and their referents is *arbitrary*: the relationship is based on an agreement that in the language of which the word is a part the word means what it does. To give an example, there is nothing cat-like in the word ‘cat’. It is simply a combination of letters that English speaking people use when they refer to a furry, domesticated, carnivorous mammal. To the contrary, images usually evoke their referents by resembling them (representational images<sup>2</sup>). This characteristic of images is often considered to enhance the efficiency of cognitive processes. The reason for this is that much of the meaning that we derive from our environment is derived perceptually. Franck Ganier (2004, 21) also points out that adding images in a user manual can reduce the cognitive load and help the user to elaborate a mental model. However, if the user has to build a mental representation from text, it will require more resources and consequently, it will induce a heavier cognitive load than that produced by images.

The second fundamental difference between words and images that Williams and Harkus present (1998, 34) is the types of referents visual and verbal mode evoke. The referents that words evoke tend to be broad and inclusive categories, whereas images usually evoke categories that are more narrow. To give an example, we can use the word “screen” to refer to a number of different kinds of screens: TV screens, computer screens, or movie-theatre screens. However, if we want to convey the concept of “screen” with an image, it would not be such an easy task. The question is: what kind of image should we choose? Image of the screen of a laptop would most probably evoke

---

<sup>2</sup> There are also images that do not realistically depict what they are intended to represent, like graphs, charts, tables, and diagrams (Williams & Harkus 1998, 36).

the concept of a computer screen rather than a screen in general. In contrast, words can evoke an entire class of elements instead of some specific referent. However, these words can be modified to convey a narrower version of the concept by using modifiers and syntactical rules. I could, for example, make the word “screen” more specific by referring to it by saying that it is “the new TV screen that my parents bought last summer”. Thus, as Collin Ware (2004, 303) points out, the greatest advantage of words over visuals is the fact that spoken and written natural language is the most elaborate and complete symbol system that we have.

In addition that words are better for broader concepts and images are better for exemplars of concepts, there is also another type of a distinction between the referents these modes usually evoke: words tend to be more efficient in evoking abstract concepts, while images work better for concrete objects. Jean-Luc Doumont (2002, 221) points out that visuals are not good at expressing abstract concepts and, moreover, they lack the accuracy that words have. In this sense, words are “worth a thousand pictures”: they can express abstract concepts unambiguously. Think, for example, of the word ‘freedom’. How would it be possible to convey this idea without using text?

However, as I already mentioned in the previous chapter, although text is often used to describe abstract concepts, adding images to accompany the text can help to clarify the abstract ideas. Lewis (1988, 242–243) states that in software documentation images can clarify abstract content, whereas in hardware documentation images can reinforce the verbal descriptions of the hardware and enhance comprehension of assembly and maintenance tasks.

The third distinction between the verbal and the visual mode in Williams and Harkus’ (1998, 34–35) model is the “differences in the structure each symbol system imposes on the information it carries. . . .” The structure of a text is linear, while images and other visual forms, like diagrams, are not constrained by the sentential structure of text. This difference is noteworthy if we think about how people store information. In the field of cognitive science, there is evidence that people store

information in hierarchical memorial structures called *schemas*.<sup>3</sup> Schemas are constructed on the basis of our experience. They organise what we already know and provide “placeholders” so that we can also organise the incoming, new information. Because of the linear structure, a text is a list of ideas and instructions that help the user to reconstitute those relationships among the ideas that the writer saw in his/her schema. Images, in turn, can preserve the view of relationships among ideas that the writer wanted to convey.

According to Ganier (2004, 21), document designers should optimise the use of the working memory, because it seems to be strongly implicated when people process instructions and also because its capacity is limited. He says that it requires more resources to build a mental representation from text than it does from images. Lewis (1988, 237) also argues that images are encoded differently in our memory than words: when we see information in image form, our perception of the features of that image interacts with our memories of real objects and with other mental images we have. That is why we can more easily remember the information we get from images.

Williams and Harkus (1998, 35) state that there are also other differences between the use of text and images that result from the fact that images are not constrained by the linear structure of text: images are often more powerful than text at representing nonlinear relationships among objects or ideas. Those relationships can be either logical (as in organization charts) or spatial (as in maps or photographs). To use Williams and Harkus' example, with the help of an image of a machine, it is easy to depict a complicated set of spatial relationships among that machine's components. A verbal description of the same spatial relationships would necessarily take the form of a list because of the sentential structure of text. Harold Booher (1975, 276) points out that images tend to be the best format for presenting locations, while text is the best format for presenting difficult series of actions. Anders Björkqvall (2009, 16) also remarks that images have good semiotic resources for

---

<sup>3</sup> For more information on schemas, see for example: Mandler, Jean M. 1984. *Stories, Scripts, and Scenes: Aspects of Schema Theory*. Hillsdale, NJ: Lawrence Erlbaum.

showing spatial relationships, and that is why images are used, for example, in maps to show where some specific building is located. On the contrary, text is the best format for describing and reflection.

The last fundamental difference between the verbal and the visual mode in Williams and Harkus' (1998, 34–35) list is “the degree to which information carried in either system can be processed perceptually.” Mostly we process the visual world rapidly and unconsciously. Much of the meaning of the visual information is understood via “pre-attentive processing”, which refers to the unconscious accumulation of information from the environment. At pre-attentive stage people do a lot of processing: the lines and boundaries are combined to reveal objects which are then separated from other objects, and from their backgrounds. In consequence, when we see an image, we do not see its individual lines unless we consciously attend to them. Naturally, pre-attentive processing happens when we look at words, too. But the difference is that meaning derived from words requires more processing – processing at the conscious level that takes place serially and that requires effort on the user's part.

Booher (1975, 266) remarks that it is easy to process information from images, and images can also present a great amount of information in a small space. In my opinion, the relative processing ease of images is an important thing to remember when document designers design documents. The fact is that people read documents because they have to: they want to learn how to use a device or they have problems using it. That is why it is important that users do not have to use too much effort and time to understand the information that the document includes.

All in all, text and images both have their advantages that result from the different characteristics of these two modes. Williams and Harkus (1998, 36) point out that the practical implication that can be drawn from the fundamental differences between text and images is that these two modes work best in concert. Words are good at expressing abstract objects and action. In addition, words are more accurate than images and words make it possible to describe things unambiguously. On the other hand, images are good at expressing concrete objects, spatial

relationships and location. They reduce the cognitive load, are remembered easier and faster than words, and they are good at comprising information. Images can also clarify abstract concepts especially in software documentation as well as reinforce the verbal descriptions in hardware documentation. Moreover, according to Lewis (1988, 241), images and other graphic illustrations provide user orientation: it is often hard for the users to find the information they are looking for in the user manual. In these kinds of situations, graphic cues can help the users to orient themselves. Images can thus also be used to draw users' attention. In addition, Lewis (1988, 244) points out that images and other visual aids can increase the motivation of the users: in general, people like images and that is why images can strengthen the users' motivation to read the manual, although images may not improve performance. So it seems that images draw attention and motivate users, because they are somehow more attractive than words. It is hard to say whether this attractiveness can be explained with the help of the fundamental features of images. However, I would argue that it is a feature that document designers should keep in mind while they create documents.

To conclude, because of their different structure, images and texts are good at expressing different kinds of things. The differences discussed in this chapter have been summed up in Table 1 below. The characteristic functions of images will be exploited later in the analysis chapters.

**Table 1: The characteristics of text and images**

<b>Text is good at:</b>	<b>Images are good at:</b>
expressing abstract concepts	expressing concrete objects
expressing difficult series of action	expressing spatial relationships/location
describing things accurately	comprising information
	orienting the user
	reinforcing verbal description
	drawing attention
	increasing motivation
	reducing cognitive load/help to build a mental model

### 3.3 Integration of Text and Images

In the previous chapter I discussed the characteristics of visual and verbal modes and listed some of their fundamental differences. However, as Fortune (2002, 105) states, this is not enough: in addition to understanding how words and images work separately, document designers need to know how they act interdependently. I have now discussed the two modes separately, and next I will move on to the relationships between them. Firstly, I will introduce different types of models that have been used to study multimodality. Secondly, I will justify my choice of model and describe it in detail.

Maier et al. (2007, 453–454) remark that document designers need to be able to exploit the meaning-making potential of multimodal communications. They state that multimodal analysis offers tools for defining which modes should be given prominence in creating different types of meanings. Maier et al. (2007) have written an article on multimodal analysis: "Multimodal Analysis: An Integrative Approach for Scientific Visualizing on the Web", which according to Aarhus School of Business (2009), has received two awards: *the New York Metro Distinguished Award* and *the Society for Technical Communication International Merit Award*. According to the authors, they attempted to establish connections between modality on the one side and document design on the other. One of the writers of the article, Constance Kampf (2009), points out that one of the reasons that made the article so respected is the timing. Since the article was published, the attention towards multimodal approach has been overwhelming in the United States. Kampf (2009) remarks that they were able to offer a valuable tool, a tool that connects multimodal theory to document design, to the target audience at a time they were looking for it. Another writer of the article, Carmen Maier (2009), states that the success of the article proves that the multimodal approach is gaining the attention that it deserves all over the world.

In the article, Maier et al. (2007) tested the multimodal approach on an interactive edutainment text aimed at multimodal literate children by using an adaptation of Theo Van

Leeuwen's (1991, 2005<sup>4</sup>) multimodal model of image-text relations. According to Maier et al. (2007, 470), van Leeuwen's model includes two types of verbal-visual relations that both have several subtypes: elaboration and extension. Elaboration means that an image provides more detail to demonstrate concepts that appear in the text, whereas extension describes a situation when an image extends or changes the meaning of the text by going beyond the verbal to make a new meaning together with the text. The subtypes of those two categories include the following:

- elaboration through specification
- elaboration through explanation
- extension through similarity
- extension through contrast
- extension through complementation (Maier et al. 2007, 464).

Maier et al. (2007, 464) state that document designers and science writers can use this categorisation scheme to base their educational decisions on functions that are derived from the interaction of the text and visuals.

The result of Maier et al.'s study is that the multimodal analysis is actually an efficient tool for selecting the suitable communicative strategies when mediating science to target groups. The study proves that the relationships that exist between the visual and the verbal mode are not only relations of co-existence but those modes interact with each other. Multimodality gives authors more opportunities to shape the audience's perceptions of the text: they can simultaneously use words and images to influence the manner in which texts are interpreted by viewers. (Maier et al. 2007, 474.)

Fei Lim (2004) has also created and tested a multi-semiotic analysis model. Lim (2004, 220) proposes the Integrative Multi-Semiotic Model (IMM) as the 'meta-model' for the analysis of pages that include both text and images, such as children's picture books and advertisements. A term "meta-model" is used to describe that the model brings together different frameworks that are now

---

<sup>4</sup> Van Leeuwen, Theo. 1991. "Conjunctive Structure in Documentary Film and Television." *Continuum* 5, 1: 76–115.  
Van Leeuwen, Theo. 2005. *Introducing Social Semiotics*. London: Routledge.

available in the field of multimodal studies. In the case of this study, one of the frameworks of the proposed model is especially interesting and relevant: *Space of Integration* (SoI). According to Lim (2004, 225), SoI can be used to study the relations between two modalities, the visual and the verbal. Lim states that when linguistic and pictorial semiotic resources interact, the total meaning is more than just adding up the meaning made by each of these independent modalities. Lim (2004, 238–239) states that SoI can be used as a theoretical platform for discussing “the dynamics in the interaction between language and visual images for meaning-making in a multi-semiotic text.” The main idea of the model is that there are two kinds of *contextualizing relations*: *co-contextualizing relations* and *re-contextualizing relations*. One of the two types of relations can always be found in a multimodal text where two modalities operate together. When two resources share co-contextualizing relations, the meaning of one modality seems to reflect the meaning of the other through some type of convergence. The resources share re-contextualizing relations when the meaning of one modality is either unrelated or even at odds with the other. So the focus of the model is on the nature of the interaction between the two semiotic modalities.

Arguably, these kinds of models that address the relationship between the verbal and the visual can offer useful information for document designers, who need to be able to integrate different modes of communication in documentation. A good example of the adaptation of multimodal thinking in document design is Schriver’s (1997) classification of the relationships between text and images in documents, which will function as an analysis model in this study. Schriver’s model seems to be widely recognised in the field of document design (see for example: Willerton 2005 and Portewig 2008). So although it would have been possible for me to use Maier et al.’s or Lim’s models, Schriver’s model was chosen because it seems to be respected and because it has the same framework as my study: the field of document design.

In her model Schriver (1997, 412–413) lists five relationships among text and images:

- Redundant
- Complementary
- Supplementary
- Juxtapositional
- Stage-Setting

The first relationship that Schriver (1997, 413) lists is *redundancy* and it means that the key ideas are repeated or paraphrased. In document design, redundancy means that similar ideas are presented in alternative representations (e.g., visually and verbally), in alternative media (e.g., paper and online), or by activating different senses (e.g., sight and sound). Redundancy can be highly useful if it is used in the right context: when it is hard for the user to fully understand a concept, the redundant relationship can be a great help. Thus, the more difficult the topic is, the more likely it is that the user will benefit from redundancy. But on the other hand, redundancy can be a nuisance if the document designer tells or shows the user something with which the user is already familiar. That is to say that the excessive use of redundancy can irritate the users and make the users think that the document designer underestimates them.

Nevertheless, although redundancy often is a good help for the users' understanding, it can be a difficult relationship for the document designer to use effectively. This is because it is often challenging to decide whether a concept is already well-known by the audience. Every audience is different, and furthermore, their background knowledge is varied.

The second relationship that Schriver (1997, 415) introduces is *complementary*. If words and images are in a complementary relationship, they employ different visual and verbal content. Both modes work together and help the user understand the same main idea. Together the two modes give a more comprehensive picture of the idea than either does alone, because each of the modes provides different information about the idea. That is to say that words and images complement each other. Schriver gives the following example on the complementary relationship:

[A] complementary text and diagram combination about how motor works might offer a 3-D presentation of the spatial features of the motor, a representation that would be cumbersome to provide in prose. On the other hand, details about how the purpose of the motor and its practical uses might be best presented in words.

Together these two modes strengthen and clarify the users' understanding of the main idea.

According to Schriver (1997, 415), the complementary relationship can also help the users to integrate the content from words and images. Each mode has a mutually constraining effect on how users understand the main idea. For instance, in newspapers a headline of an article may guide the user to interpret a photo on the first page in a certain way. In short, Schriver (1997, 418) states that, when words and images are in complementary relationship, they complement each other because each mode provides essential information that the other mode does not provide. Consequently, this helps the user to understand the distinction.

Schriver (1997, 417) mentions that when words and images are in a complementary relationship they can provide complete information about the action to take: the images give the user spatial cues about where to press or pull, while the text offers exact information about what to do and when to do it. These kinds of complementary text and image relationships might be useful, for example, in procedural instructions that *Lumia 800*'s user manual presumably includes. User manuals of mobile phones often include step-by-step instructions which require both text and images to be effective and understandable, for instance, instructions for inserting the SIM card. In these kinds of instructions both text and images provide essential information that is not provided by the other mode: the image shows **where** and the text describes **how**.

According to Schriver (1997, 418–419), words and images can also be arranged so that one mode is the dominant one, providing most of the content, while the other one supports and elaborates the points that the dominant mode makes. This kind of relationship is called *supplementary*. Schriver (1997, 419–420) states that when words and images supplement each other, they often occur in the form of examples: an image may illustrate something that is hard to understand only with words, or a sidebar may unpack an image. If the user has trouble imagining

what is intended, supplementary words and images can help to clarify the content or expand the ways in which the user interprets the main ideas.

Schrivier (1997, 421) advises document designers to plan carefully how they make the supplementary combination of words and images to function within the structure of a document. She points out that unneeded additions can distract the user and unsystematic additions can make the user confused. If document designers add images randomly, they may inappropriately lead the user to believe that topics that include images are somehow more essential than those that do not.

The fourth relationship Schriver (1997) introduces is called *juxtapositional*. According to Schriver (1997, 422), when text and images interact through a juxtapositional relationship, “the main idea is created by a clash, an unexpected synthesis, or a tension between what is represented in each mode.” Users cannot understand the intended idea unless they see both text and images simultaneously. Schriver (1997, 424) states that juxtapositional relationships are most often used in advertising, design, poster art and cartoons, for example.

The final manner in which words and images can interact is through a *stage-setting* relationship. According to Schriver (1997, 424–425), in a stage-setting relationship “one mode provides a context for the other mode by forecasting its content or soon-to-be presented themes.” The aim is to help the users to get a sense of the big picture before they begin. In document design these kinds of stage-setting relationships can be useful at the beginning of chapters in multi-chaptered documents. An image can be conjoined with the title of the chapter, for example, and this can give the user a feel for the theme of the content. However, the stage-setting relationship can also do more than just provide a visual anchor: sometimes the purpose of this relationship is to shape the users’ attitude about the content in some particular way. For example, a drawing of a child using a mobile phone might be used at the beginning of the user manual to convey the idea that the phone is easy to use.

Schrivier (1997, 424) points out that document designers tend to be a bit conservative in their image-text combinations, the most common relationship being the supplementary relationship. I

would argue that this excessive use of supplementary relationships arises from the assumption that in the instruction materials visuals are often presented as subordinate to text. Ware (2004, 315), for instance, states that visual and verbal languages are not on equal footing with each other: we are all experts in verbal language, having been trained at it from an early age, but we are not experts in visual communication. Because of this dominance of words as a medium of communication, visualisations are in Ware's words "hybrids" which are used only where there is a clear advantage of using them. This shows that the verbal mode is still considered to be the mode on which document designers should focus, whereas images are used "only when needed". However, if images are always seen merely as supporters, how can document designers make effective choices when creating multimodal documents? In my opinion, instead of thinking "where should I use images to support my text", document designers should think "how should I integrate text and images to convey the idea as clearly as possible".

In Schriver's (1997, 424) opinion there is much more room for creativity in image and word combinations in document design. However, I think that without extensive guidelines for text-image integration it is unlikely that document designers use their creativity and try different kinds of ways to integrate text and images. Portewig (2004, 32) also points out that in order to address our problems with combining visual and verbal information, we need a framework that deals with the effects and importance of the visual information in document design.

In my opinion, Schriver's five basic ways of combining text and images just presented above can be useful for document designers. With the help of the model, document designers can become more consciously aware of the choices they make when they integrate text and images. However, I believe that the model is not alone sufficient to help document designers to make decisions about effective integration of text and images, because, as mentioned in chapter 2.1, document designers are not often experts at using images. That is why this study focuses both on the characteristics images and on the relationships they form with the text. In order to successfully integrate text and images, document designers need to be aware of how different modes work alone and together.

## 4 Material and Methods

In this chapter, I will more thoroughly present the material of this study and the methods that are going to be exploited in order to get the results. Firstly, I am going to provide basic information on the two user manuals on which this study focuses. I will introduce the two products, *Lumia 800* and *Gemini*, and provide such information on the products' user manuals that is relevant for the analysis of text and images. Secondly, I will describe how I am going to conduct the analysis of the material.

### 4.1 Lumia 800 and Gemini

*Lumia 800* is the first smartphone by Nokia on the Windows Mobile platform. According to Matti Pulkkinen (2011), who is a Senior Consultant of Technical Writing at Lionbridge, it is a high-profile product targeted for *early adopters* of new technology<sup>5</sup>. The main target group are males of 25 to 45 years of age who are interested in technology and in new innovative products. Thus, in my opinion, the user manual of the phone should also respond mainly to the needs of the kinds of users who presumably are familiar with the kind of technology that *Lumia 800* represents, or if they are not, at least they are motivated to learn to use it.

The user manual of *Lumia 800* is available through Nokia's webpages. The manual is in PDF form, which basically means that the users have three ways to search for the information they need: they can scroll up and down the pages, they can use the hyperlinked table of contents or they can use the search function to find what they are looking for.

The user manual includes in total 32 images that will all be analysed. All of those images are line drawings, coloured with blue and orange. There are two other types of images in the user manual as well, namely icons and warning images, but those will be left out of the study. The reason for this is that both icons and warning signs represent rather special type of images. Laura Gurak (2003/1992, 492) defines an *icon* as "a symbol that represents, either alone or in combination

---

<sup>5</sup> The term originates from Rogers, Everett M. 1962. *Diffusion of Innovations*. New York: Free Press.

with some text, a task that users will perform.” In the user manual of *Lumia 800*, an example of this kind of visualisation is the icon for making a call (an image of a phone). According to Gurak (2003/1992, 492), users recognise icons and associate them with something with which they are familiar. In other words, when people see icons they do not learn new abstract concepts but, rather, they connect the icon with something they already know. In my opinion, this characteristic of icons clearly makes them a unique group of images that would require a study of their own. Moreover, it is reasonable to exclude icons from the analysis, because document designers are not the ones who decide how the icons should look. Naturally, icons in the manual look exactly as they look in the actual device. In addition, warning images are excluded from the study because their usage is highly regulated and standardised, and that is why their usage should also require a separate study.

*Gemini* is Lionbridge’s internal project management utility that is interlinked with other internal tools of Lionbridge. *Gemini* has three distinctive user groups: Operations (project managers and senior project managers), which uses *Gemini* to forecast projects and to manage project work; Finance, which uses *Gemini* to control invoicing, credit notes and reporting, for example; and Vendor Management, which uses *Gemini* to control vendors. So different user groups use *Gemini* for different purposes. However, I would claim that *Gemini*’s users are quite a homogeneous group in that sense that they can all be considered to be professionals in their own field.

The user manual of *Gemini* is a Microsoft Compiled HTML Help, which is Microsoft’s proprietary online help format. The user manual consists of HTML files, a hyperlinked table of contents, and an index file: in order to get the information, users need to exploit either the table of contents or the index list that displays an alphabetic list of search terms. By clicking the hyperlinks in the table of contents or in the index list, the users can open the HTML files they want to read. In addition, the users can navigate in the document by using the search function of the help.

The main types of images used in *Gemini* user manual are screen captures. Previously the documentation also contained information about processes, illustrated with simple flowcharts. However, nowadays descriptive content and flowcharts are avoided. The reason for this is the

notion that the end-user uses documentation to solve problems, not to learn about product concepts. In the manual, every section is labelled with symbols that represent different user groups. This makes it easy for the different user groups to identify the information that concerns them.

(Pulkkinen 2011.)

The total amount of screen captures in the user manual of *Gemini* is 65. The only images that are not screen captures are the symbols for different user groups. I have excluded these symbols from my study, because they do not have an actual connection with the text. Their only purpose is to make it easier for the users to find the information that concerns their area of expertise.

## 4.2 Method

The aim of this chapter is to describe the method used in this study as comprehensively as possible. The method is two-fold: I will begin by analysing the visual elements in the two user manuals, and after that, I will focus on the relationships that exist between text and images. In the latter part of the analysis, I will see whether there are some connections between the functions of the images and the relationships they form with the text.

### 4.2.1 Images

In this first phase of the analysis, my intention is to find out what the function of images in the manuals is. I am going to use an Excel table to categorise the function of each image. With the help of the table it is easy to see how many of the images are categorised to have the same function. So the main focus will be on the images, but deciding their functions will, of course, also require a holistic inspection of both images and text.

As a theoretical background for the analysis of *Lumia 800*, I will take advantage of the table of the characteristics of text and images that I presented in chapter 3.2: in the case of each image, I will analyse whether the function of the image is to express spatial relationships, help to locate information, or perhaps reinforce abstract concepts, for example. When analysing the functions of

images in the user manual of *Gemini*, I will take advantage of theoretical studies that focus specifically on screen captures: the user manual will be analysed using Hans van der Meij and Mark Gellevis's (1998) model on the roles of screen captures, which will be presented more thoroughly in chapter 5.1.2. However, the roles will be complemented with elements from the same model that is used to analyse *Lumia 800* where appropriate.

#### **4.2.2 Relationships between Text and Images**

In the second part of the analysis, I will conduct a contrastive analysis of text and images with the help of Schriver's (1997) model that was introduced in chapter 3.3. The aim is to find out, whether the five relationships – redundant, complementary, supplementary, juxtapositional and stage-setting – appear in the materials.

I will analyse the relationships that exist between text and images and enter the results in the same table in which I entered the functions of the images. Having one table will allow me to see potential connections between the functions of the images and the relationships that they form with the text. It would be interesting to see whether images that express spatial relationships are always integrated similarly with the text, for example. In addition to this comparing inside the user manuals, my intention is also to do comparisons between them. I hope to find some differences that arise from the fact that these two documents represent different user manual types: one is a user manual for a concrete object and the other is a user manual for software. By analysing and comparing the use of images and the relationships they make with the text, I wish to find guidelines that document designers can use when they visualise documents.

## 5 Analysis

In this chapter I am going to introduce the results of the analysis of the two user manuals. The first part of the analysis focuses on the usage of the images in the documents, whereas in the second part the focus is on the relationships between text and images.

### 5.1 Use of Images in the Manuals

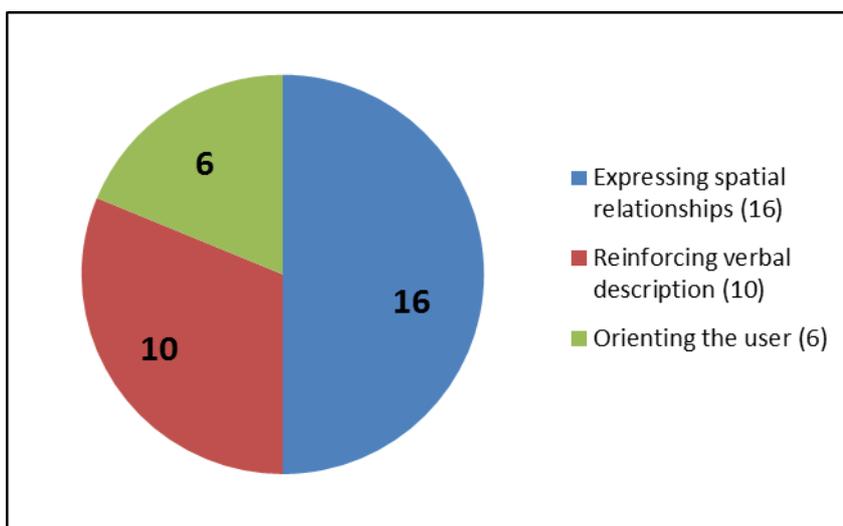
The aim of this chapter is to provide information on the functions of the images in the user manuals of *Lumia 800* and *Gemini* and to reflect on the similarities and differences between the two user manuals. What is important to note before starting the categorisation of the images is the fact that absolute categorisation is always complicated, if not impossible. One could say, for example, that helping to locate information is a purpose that every image in different kinds of user manuals fulfils. My purpose is to interpret, with the help of the context in which the image occurs, what the most **significant** function of each image is. When reporting the analysis, I will give examples which clarify the grounds on which I have analysed the functions of the images. As already mentioned, I will take advantage of Table 1 on the characteristics of text and images on page 27 to guide my analysis, focusing on the features of images. According to the table, images are good at:

- expressing concrete objects
- expressing spatial relationships/location
- comprising information
- orienting the user
- reinforcing verbal description
- drawing attention
- increasing motivation
- reducing cognitive load/helping to build a mental model

### 5.1.1 Lumia 800

There are in total 32 images and 88 pages in the user manual of *Lumia 800*. The images are not spread evenly throughout the user manual: most of the images occur in the first quarter of manual. I believe the reason for this is that most of the images at the beginning of the user manual help the user to deal with the concrete features of the phone: locating the different parts of the phone, learning to use the touch screen, learning how to insert SIM card and how to charge the phone, for instance. Towards the end of the user manual, the instructions concentrate on more abstract features of the phone, such as using the internet and Bluetooth. As images are often better for describing concrete objects, whereas text works better for abstract concepts, the concentration of images at the beginning of the user manual is quite natural. However, I wonder how this affects users' motivation. Are users motivated enough to read the last three quarters of the text that do not include so many images? This issue will be discussed more thoroughly later in this chapter.

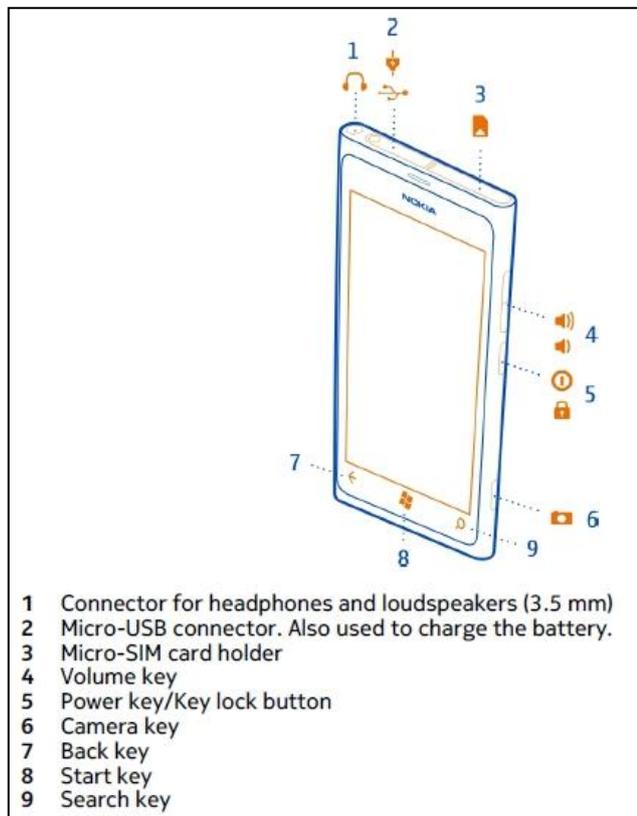
Based on my analysis, there are in total three different types of primary functions that the images have in the user manual of *Lumia 800*. In summary, there are images that express spatial relationships, images that reinforce verbal description, and images that orient the user. Figure 2 depicts the number of each function and the distribution between the three functions:



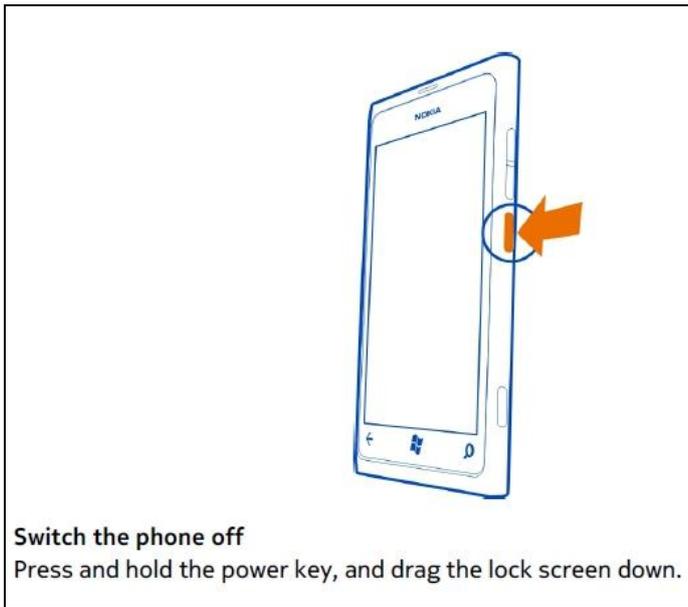
**Figure 2: Functions of images in *Lumia 800*'s user manual**

As the figure shows, images are most commonly used to express spatial relationships. This kind of frequent usage of images for spatial relationships is not surprising. As Williams & Harkus's (1998, 34) list of differences between text and images indicates, those two modes differ in the nature of the referents they evoke (difference number two in chapter 3.2). Thus, the most obvious explanation for the high number of images that express spatial relationships is the fact that text tends to be more efficient in evoking abstract concepts, while images are often better at expressing concrete objects (see for example Doumont 2002, 220–221). The user manual of *Lumia 800* gives instructions about a concrete object, a phone, and that is presumably why expressing spatial relationships is so common. Examples 1 and 2 offer good examples of how images can be used effectively to help users to locate different parts of a concrete object:

**Example 1: Image that expresses spatial relationships 1 (*Lumia 800*, page 6)**



**Example 2: Image that expresses spatial relationships 2 (*Lumia 800*, page 13)**



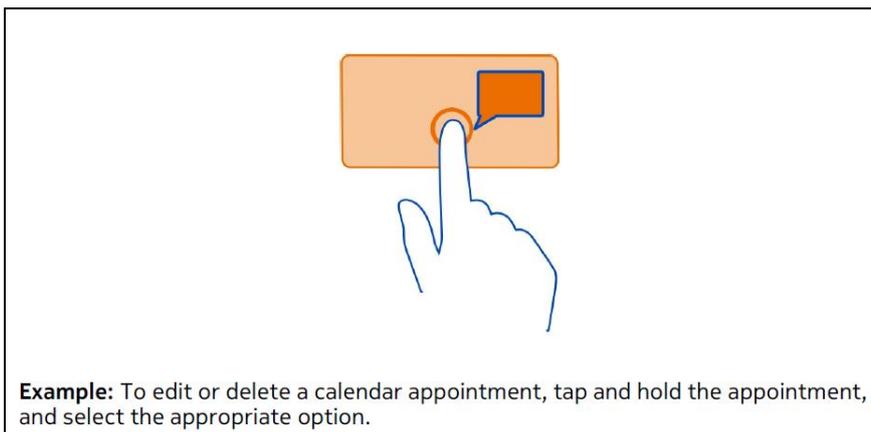
By inspecting Examples 1 and 2, it is easy to realise how superior images are to text when it comes to locating specific parts of objects. It would be extremely complicated to describe different buttons and other parts of the phone by using only text. An image can tell in a glance something that is very cumbersome to describe with words. To give an example, if a document designer tried to explain the location of the power key of *Lumia 800* without using images, it would probably look something like this: “The **power key** is the oval button on the right side of the phone. It is between the **volume key** and the **camera key**.” The obvious next question is: how does the user know what **volume key** and **camera key** are?

As it is evident from the example above, it is possible to express spatial relationships with words. However, it is considerably more laborious for the user to process the information from the text than it is from the images. As discussed in chapter 3.2, building a mental model from text requires more resources than it does from images. In addition, trying to explain all the spatial information in words would take much more space than it does with using images, because visuals are good at comprising information. Nowadays, in the era of the internet, it is important to make user manuals compact, because people are used to acquiring knowledge fast. If people feel that

there is too much to read in the document, they probably do not read at all (see for example Schriver 1997).

As I have already mentioned, images are good at expressing concrete objects while text is better for abstract objects and action. Nevertheless, sometimes text is not enough to express abstract ideas or difficult series of actions and that is why images are needed to support the ideas or actions that are presented in the text. As Figure 2 shows, these types of images are the second most common in the user manual of *Lumia 800*. All of the images that reinforce verbal description in the manual help the users to understand actions that they need to complete. A good example of an image that reinforces verbal description of action can be seen in Example 3:

**Example 3: Image that reinforces verbal description (*Lumia 800*, page 20)**

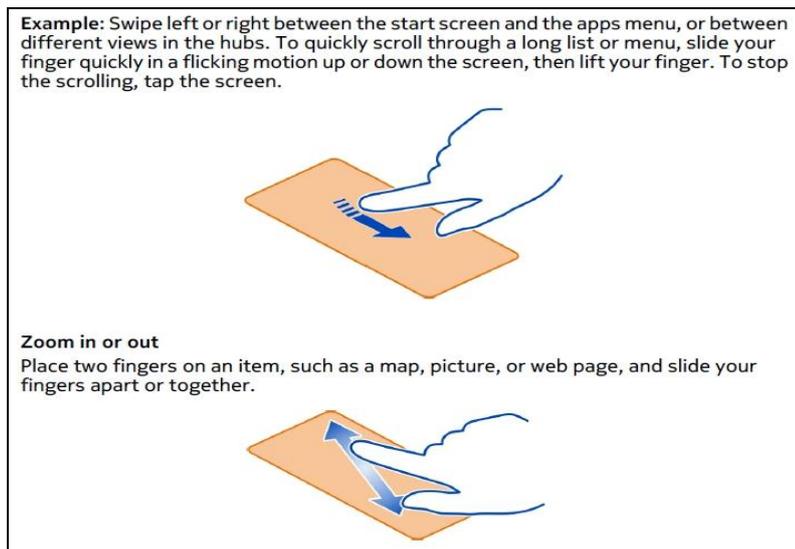


Although the verbal description of the action can be said to be a comprehensive one, the image helps the user to understand what the text conveys. These kinds of images are normal in hardware documentation, as pointed out in chapter 3.1: in hardware documentation, images can reinforce the verbal descriptions and enhance the users' comprehension. Images as the one in Example 3, go under this category. I would not classify them to be images that express spatial relationships, because the most accurate spatial clues are clearly given in the text: "tap and hold the appointment". The image does not depict an actual calendar appointment that the user should tap. Instead, it clarifies that by pressing the appointment long enough, a context menu (a speech bubble)

appears on the screen and makes it possible for the user to continue. In the user manual, all these types of images appear in connection with instructions that advise the users to use the touch screen. This is presumably a good choice, because touch screens are a rather new technology. In addition, although the users of *Lumia 800* are supposed to be interested in new technology, it does not mean that they know much about touch screens.

The reason why images are especially important is the specialised vocabulary that is used when explaining the touch screen features. In connection with the images that are classified as reinforcing verbal description, the following verbs that describe the handling of the touch screen often occur: *tap*, *slide*, *swipe*, *zoom* and *scroll*. In Example 4 there are two example images that appear together with these verbs:

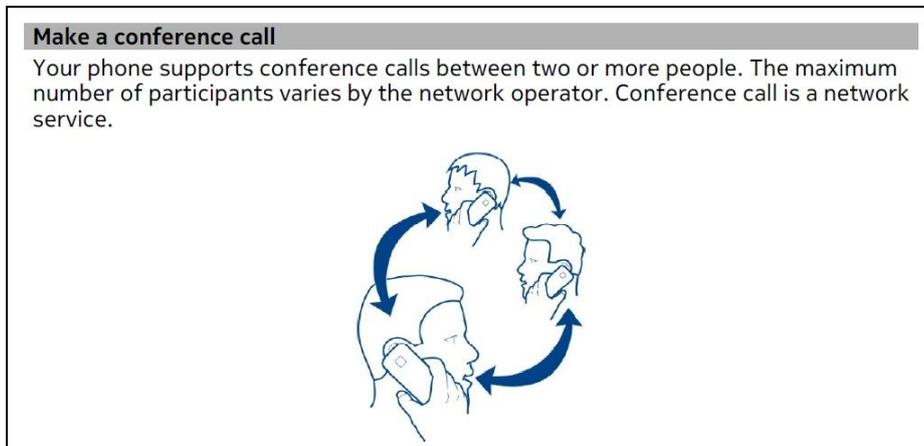
**Example 4: Instructions for swiping and zooming (*Lumia 800*, page 21)**



I would claim that for a person who does not have experience in touch screen features, these verbs are not so straightforward to understand. That is why images that accompany these verbs offer valuable information that will certainly help the users to understand the instructions better.

Finally, as Figure 2 shows, there is also a third type of function of images in the user manual: images that orient the user. An example of an image that falls under this category can be seen in Example 5:

**Example 5: Orienting image (*Lumia 800*, page 31)**



The image appears under the title “Make a conference call” and it depicts three people who are talking on the phone. The pointing arrows between the heads of these people make it easy to understand that all these three people are connected via phones.

The interpretation as to which category best describes these types of images was a rather complicated task. At first, it seemed that these images do not have an actual function at all, because they do not add anything to the text that accompanies them. Rather, they just visualise the action about which the text gives instructions. That is why I first categorised these images as purely decorative images, but upon closer examination I realised that these images are in fact “orienting” the users. Practically, these images help the users to find the information they are looking for. As pointed out in chapter 3.2, it is often hard for the users to find the information, and in these kinds of situations images can help the users to orient themselves. Ganier (2009, 403–404) also notes that by providing visual indication, such as images, it is easier for the users to search and locate information.

Using images to orient the user is related to the way in which people read user manuals in general. According to Ganier (2009, 410), procedural documents are rarely used before using the device. Rather, people use procedural documents to search for answers to particular problems that occur when people use the device. Ganier's (2009) study, in which he inspects how people deal with procedural documents when using a new device, shows that when people search for information, they often use images to find the information they are looking for and then the accompanying text to perform the task. A conclusion that Ganier (2009, 410) draws is that the mixed format (text + images) helps the users to find the information within the user manual more quickly. Consequently, orienting images have an important function in helping the users to locate information faster.

What is important to consider here, is whether images that orient the user are equally useful in all kinds of user manuals. In printed user manuals, these images are crucial to finding information fast and effortlessly, because the only way to find information is to use the table of contents, the index list or browse through the pages manually. On the other hand, in PDF user manuals, such as *Lumia 800*, in addition to scrolling up and down the pages, users can search for information with the help of the search engine or the hyperlinked table of contents. Because of these electronic search methods, it is probably easier for the users to find information they are looking for. However, PDF manuals also consist of separate pages that the users need to browse if they are unable to find the information with the help of the search engine or the hyperlinked table of contents. In these kinds of situations, the images that orient the users to find the information may become highly useful.

Jakob Nielsen (2000a), a leading web usability professional, offers interesting findings on his websites about navigation in an electronic environment. His studies show that when users scan a new page they look straight at the content and ignore the navigation areas. According to Nielsen, people rarely read carefully online, but instead they scan the page. Judging from these comments, it seems that people do not use navigation tools as effectively as they could. I believe that the same kind of scanning of the contents that people do when they read websites happens most probably

with electronic documents, too. Thus, a conclusion can be drawn that orienting images can, in fact, also be useful in electronic documents.

Another interesting point that Nielsen (2000b) brings forward when he discusses the web browsing habits of people is the superiority of text to attract attention. In general, users do not look at the images first, but instead they are drawn to headlines, article summaries, and captions. This comment seems to be at odds with my previous comments on the orienting effect of images in documents. However, I believe there is an obvious reason for this: when people browse websites, they often see different kinds of advertisements all over the pages. Advertisements often include eye-catching images the purpose of which is to draw users' attention. Nielsen (2000a) points out that people are accustomed to ignoring advertising on websites and anything that looks like an advertisement. This is presumably the reason why people mostly rely on text when they look for information on the internet. However, electronic documents are a source of information that does not usually include advertisements, and thus, I would say that people have a different attitude to start with. As stated in chapter 3.1, images in documents often create a feeling that the information is important, and so images in electronic documents can have an important function in grabbing the users' attention and helping them to locate information. However, it is important to notice that HTML manuals, such as the user manual of *Gemini*, requires the use of the navigation tools: the users need to use the hyperlinks to open the files they want to read. Nevertheless, I suppose that orienting images in HTML manuals also help the users to locate information inside the file they are reading. Overall, however, proving the usefulness of the orienting images in electronic user manuals would require conducting a usability test, which is beyond the scope of this study.

In addition to using orienting images to guide the users, I would recommend them for another purpose as well: to motivate the users. As I mentioned at the beginning of this chapter, there are not so many cases towards the end of the user manual that would actually require images. However, I would claim that images are needed at the end of the user manual as well to motivate the users. Nicole Loorbach, Joyce Karreman and Michaël Steehouder (2007) make interesting conclusions

about the effects of adding motivational elements to user manuals: they found out that motivational elements did not always affect the users' efficiency to use the device but they increased the users' satisfaction with the manual. Arguably, this increase in satisfaction may in turn motivate the users to keep on trying when they are on the verge of giving up with the manual.

Although Loorbach et al. (2007) focus only on the motivating effects of text, it is still reasonable to assume that images also have the same kind of effects on the users. They list, for instance, colours as motivational elements that catch the users' attention. In addition, I would claim that in the same way as colours, images can certainly be used to draw users' attention as well. In the user manual of *Lumia 800*, orienting images work well for this kind of motivational purpose: it is easy to add some orienting images at the beginning of new chapters, for example. This could presumably increase the users' satisfaction with the manual and keep them using the document longer. However, proving this claim would again require a usability test.

To conclude, it is not surprising that images in the user guide of *Lumia 800* are most often used to express spatial relationships. A phone is a concrete object, the usage of which requires locating different buttons and other parts of the device. On the other hand, the uses of images that reinforce textual description are also frequent. Naturally, there are always things that are not easily described with using text-only format and that is when reinforcing images can be very useful. In the user manual of *Lumia 800*, the instructions that guide the use of the touch screen clearly need the supporting images to clarify the message of the text. Finally, there are also images that help the users to locate the information. These kinds of orienting images have an important function in the user manual in guiding the users to find the information they are looking for.

### **5.1.2 Gemini**

As mentioned in chapter 4.1, the analysed images in the user manual of *Gemini* are all screen captures. Screen captures are a rather special type of images and that is why it is reasonable to

introduce their characteristics before discussing the results of the analysis. The following articles form a background for the analysis of the images in the user manual of *Gemini*.

Van der Meij and Gellevij (1998, 2002, 2004) have studied the use of screen captures in software documentation a great deal. Van der Meij and Gellevij (1998, 529) point out that screen captures have not received much attention in the literature on document design, which is rather surprising because screen captures are probably the most frequently used images in software manuals. For example in Finland, most of the document designers work in the field of information technology (STVY<sup>6</sup>). In their article van der Meij and Gellevij (1998, 529) present four main roles of screen captures. They propose that screen captures can help users to:

- Switch attention
- Develop a mental model of the program
- Verify screen states
- Identify and locate window elements and objects.

The first role refers to three distinct sources of information to which the users must pay attention when reading software manuals: an input device (keyboard, mouse), a manual, and a screen. When reading software manuals users must regularly switch attention to and from these three sources of information. Screen captures can help the users with this switching by prompting the user to attend the screen at the right moment and by providing a clear point for reentry into the manual after attending the screen. Secondly, screen captures help users to form a mental model of the program. Screen captures can familiarise users with the main windows, explain them the spatial layout of the window, or develop a sense of logical flow of windows. The third role of the screen captures is to verify screen states. Images support the users' progress and they facilitate random access entry into the manual. That is to say that screen captures make it easy for users to verify whether they have done the right thing and they also catch the users' eye and help them to find information for a specific problem they have encountered, for example. The last role mentioned in

---

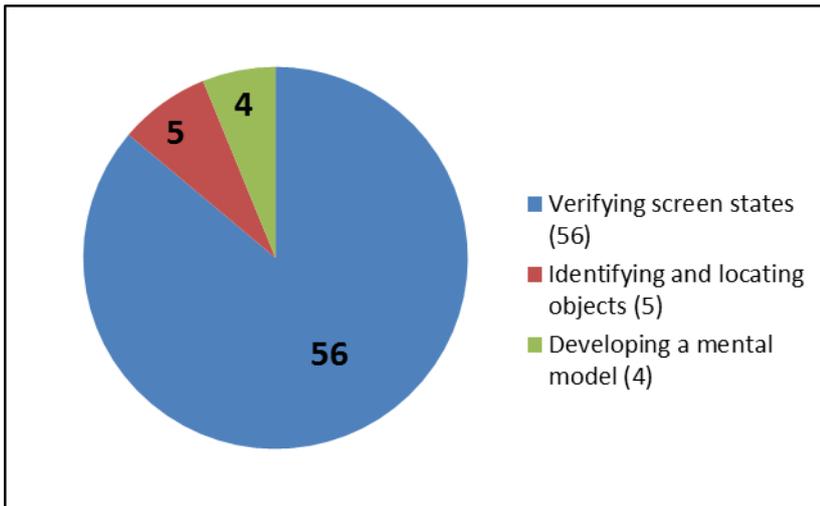
<sup>6</sup> Suomen teknisen viestinnän yhdistys (the Finnish Technical Communications Society)

van der Meij and Gellevij's list is the supporting effect of screen captures on the identifying and locating of window elements and objects. By focusing users' attention on the relevant part of the window, screen captures help users to identify and locate different objects that appear on the screen. (van der Meij & Gellevij 1998, 530–538.)

In another article, Gellevij and van der Meij (2004) report on empirical studies that were carried out to prove that screen captures in fact serve the four main functions presented in the previous article (1998). The overall conclusion that Gellevij and van der Meij (2004) draw is that the empirical studies validate the framework of screen captures functions. They found evidence that when screen captures support specific functions, the users benefit. In the following, *Gemini*'s user guide will be analysed using van der Meij and Gellevij's (1998) four roles of screen captures. However, the roles will be complemented with elements from the same model that was used to analyse *Lumia 800* where appropriate.

In my opinion, literature that focuses specifically on screen captures is needed, because screen captures may have some functions that other types of images do not have. Van der Meij and Gellevij (1998, 529) also make an interesting remark by pointing out that in Schriver's (1997) book on document design screen captures are not discussed at all, although this book broadly discusses the principles on visual design and interplay between text and images. As Schriver's (1997) book functions as an important source of information in this study, it is reasonable to support the ideas presented in the book with articles by van der Meij and Gellevij (1998, 2002, 2004).

Based on my analysis, most of the images in the user manual of *Gemini* can be classified as what van der Meij and Gellevij (1998) call screen captures that verify screen states. As Figure 3 on the next page shows, the number of screen captures that verify screen states clearly outnumber other types of images:



**Figure 3: Functions of images in *Gemini*'s user manual**

Van der Meij and Gellevis (1998, 536) point out that screen captures verify screen states both by supporting progress checks and by helping the users to quickly find the answer to a particular problem. An example of a verifying image can be seen in Example 6:

**Example 6: Image that verifies screen state (*Gemini*)**

**To Add a New User**  
 The Superuser handles this task when a new employee needs access to Gemini.

- Go to **Maintenance > User > Add a New User**.  
 The Add a new User window opens.

- Define the user details. Note that:
  - The **User Login Name** field must contain the full name of the employee in the following format:  
 First name <space> Last name  
 For example: John Doe

In Example 6, when users notice that the screen state matches with the screen capture they know that they have done the right thing. After the verification, users probably feel confident to move on with the procedure. In these kinds of situations the function of the image seems to be to motivate the user to keep up with reading and using the software. However, van der Meij and Gellevis (2002, 337) have proven this to be false: screen captures do not have a statistically significant effect on the motivation of the users. Thus, it can be concluded that the function of these images is not to motivate the users, but to draw their attention and prompt them to look at the screen and verify their progress at the right moment. Although this verification does not increase users' motivation, at least it ensures that the users stay on the right track.

Furthermore, these kinds of images are certainly useful in orienting the users as well: if the users are looking for instructions for a specific function, for example adding new users as in Example 6, the image functions as a visual anchor which catches the users' eye. According to van der Meij and Gellevis (1998, 536), "[a] quick scan can reveal whether the screen capture is what the user is looking for." Thus, by catching the users attention, screen captures that verify screen states both confirm the steps that the users are taking and orient them to find the information they are looking for.

Actually, screen captures' capability to verify screen states is the only function that could not be matched up with my model that was used to analyse *Lumia 800*'s user guide. This function can be said to be unique to screen captures and that is why it is not surprising that there is no equivalent function in my list. Nevertheless, screen captures that verify screen states seem to have some characteristics in common with images in general: they catch the users' eye and help them to orient themselves.

Although screen captures that verify screen states clearly are the most dominant types of images in the user manual, there are also two other types of images. To begin with, there are images that help users to identify and locate information. According to Meij and Gellevis (1998, 538)

screen captures can help the users to focus their attention on the relevant part of the window. An example of this kind of image can be seen in Example 7:

#### Example 7: Locating screen capture 1 (*Gemini*)

**To Add Data to the Dashboard**

When you are logged in Gemini.NET, you can add data to the dashboard in the form of DataWindows.

- In the **Select a report and add it onto DataWindow panel** dropdown menu, select the data you want to view on your dashboard.  
Currently, you can select to view active customers or projects with invoicing opportunities:
  - The **Active Customer DataWindow** shows a list of customers assigned to your office. The data includes contact details, CODA and tax registration codes, payment terms, tax bands, and customer currency.



Customer ID	Customer N.	Contact Address 1	Contact Address 2	Contact Address 3	Contact Address 4	Contact Address 5
30406	Microsoft Informa...	EPDC 4 Sandyford	Dublin 18 Ireland			
30406	Microsoft Informa...	EPDC 4 Sandyford	Dublin 18 Ireland			
30406	Microsoft Informa...	EPDC 4 Sandyford	Dublin 18 Ireland			
30406	Microsoft Ireland	Sandyford Industrial Est...	EPDC 2	Dublin Ireland		
30400	Microsoft Ireland	Sandyford Industrial Est...	EPDC 2	Dublin Ireland		
30400	Microsoft Ireland	Sandyford Industrial Est...	EPDC 2	Dublin Ireland		
7476	Microsoft Ireland ...					
7476	Microsoft Ireland ...					
7476	Microsoft Ireland ...	Account Payable Depart...	PO Box 8337	Foxrock	Dublin 18	Ireland
73624	Microsoft Ireland	Microsoft Account Receiv...	PO Box 8	Cambridge	United Kingdom	

The text describes the “**Active Customer DataWindow**” and the image shows the users what the window is like. Thus, the manual makes it easier for the user to locate the relevant window in the software. Another situation when text often needs an image to complement it is when there are a lot of objects in one window. According to van der Meij and Gellevij (1998, 538), screen captures that help the user to locate objects are especially important when windows are crowded with elements: screen captures can both reduce errors and speed up the completion of the task. In Example 8 on the next page, there are several text fields, drop-down menus and buttons on the screen capture. The text advises the users to “Change the **Project Status** to **Completed**” and the image helps them to locate the “Project status” field by emphasising the word “Completed”:

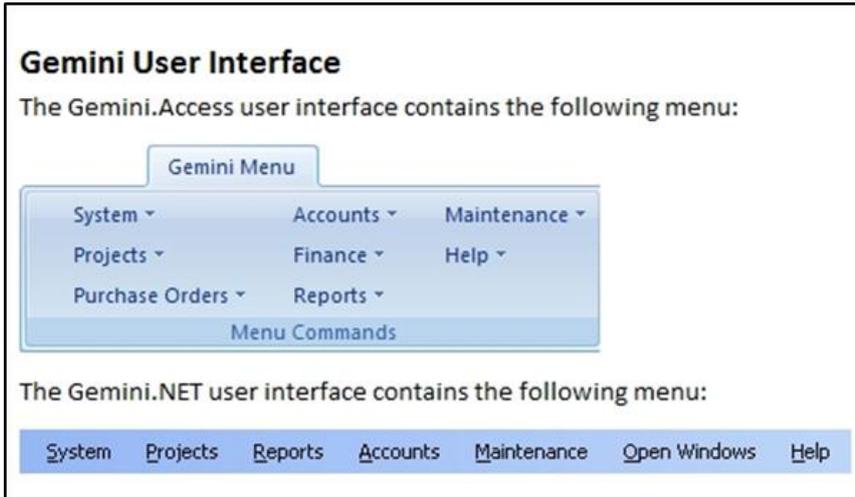
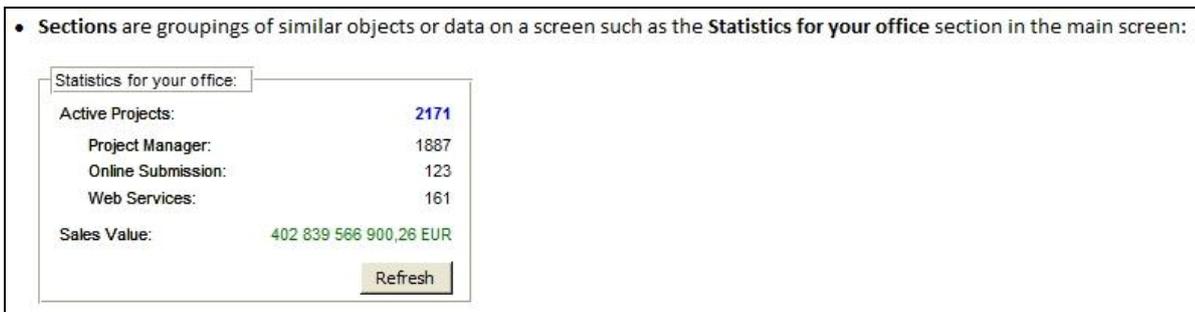
### Example 8: Locating screen capture 2 (*Gemini*)

2. To select the project, click the **Project Picker** (the .. button) or enter the project number and click **Next....**

3. Change the **Project Status** to **Completed**.

These kinds of screen captures are easy to fit into the model that was used when analysing the user manual of *Lumia 800*. Considering the model, these images could clearly be classified as images that express spatial relationships and location. That is to say that screen captures, in fact, share features that are characteristic of images in general: images help the users to locate information.

As Figure 3 on page 51 shows, images in the manual are also used to help the users to develop a mental model. Two examples on these types of images can be seen in Examples 9 and 10. Both of the images appear in connection with the introduction of *Gemini*'s user interfaces. The function of the images is to make the user familiar with the main elements of the software:

**Example 9: Screen capture that helps to build a mental model 1 (Gemini)****Example 10: Screen capture that helps to build a mental model 2 (Gemini)**

According to van der Meij and Gellevij (1998, 532), screen captures can help the user to build a mental model by acquainting the user with the main windows or explaining the spatial layout of the window, for example. That is to say that these kinds of screen captures reduce the cognitive load of the users: as mentioned in chapter 3.2, if users have to build a mental representation from text it will induce a heavier cognitive load than that produced by images.

According to van der Meij and Gellevij (1998, 532), it is important to help users to develop an understanding of the spatial layout of screens, because without knowledge about the design of the screens, the users would constantly be overwhelmed by the information they see on the screen.

Such images as Examples 9 and 10 familiarise the users with the basic parts of the software and gives them a general idea of the structure of the software. This kind of information is presumably

important at the beginning of the user manual of *Gemini*, as well as in other software user manuals: by knowing the basic structure before starting to use the software helps the users to build a mental model and use the software more effectively.

This kind of use of images can also be found in the list that was used in the analysis of *Lumia 800*'s user manual. According to the list, images are good at reducing cognitive load and helping to build a mental model. This means that my list of the characteristics of images and van der Meij and Gellevij's (1998) list actually share some functions: screen captures' capability to identify and locate objects and to help people to form a mental model are features that appear both in my and in van der Meij and Gellevij's (1998) lists.

All in all, most of the images used in *Gemini*'s user manual are images that verify screen states. These kinds of images are characteristic of software user manuals and they have an important function in keeping the users on the right track and orienting them to find what they are looking for. The user manual also includes images that help the users to locate and identify objects on the screen and help them to create a mental model of the program.

If compared with the images used in the user manual of *Lumia 800*, it is clear that software and hardware documentation differ in their use of images. The functions used in the studied manuals correlate well with the characteristic features of software and hardware documentation presented in chapter 3.1. Images in hardware documentation are especially useful in representing equipment, systems, and components. Images also reinforce the verbal descriptions of the hardware documentation. On the contrary, in software documentation, an important function of images is to describe conceptual processes and procedures, because images clarify abstract content. They enhance understanding and help to remember the information.

In short, my analysis shows that images in the two studied user manuals are in fact used for different purposes: images in *Lumia 800*'s user manual focus on describing concrete objects, reinforcing the verbal descriptions and orienting the users. Images in the user manual of *Gemini*, on the other hand, focus on clarifying abstract content by describing the structure of the software and

by verifying that the users keep up with the instructions. In addition, images in *Gemini*'s user manual also help the users to locate different objects in the window.

## 5.2 Integration of Text and Images in the Manuals

The aim of this chapter is to provide information on how text and images are integrated in the two studied user manuals. Firstly, the focus will be on the relationships between text and images in the user manuals of *Lumia 800* and *Gemini*. In addition, the aim of this chapter is to find out whether there are some connections between the functions of the images and the relationships they form with the text. What I am hoping to achieve is that I can draw such conclusions on the interplay between text and images that can function as general guidelines for document designers when they visualise documents.

### 5.2.1 Lumia 800

As mentioned in chapter 4.2.2, the analysis of the relationships between text and images was conducted with the help of Schriver's (1997) integration model. Based on my analysis, in the user manual of *Lumia 800*, there are four relationships between the written and the visual mode. Those relationships are all presented in Figure 4:

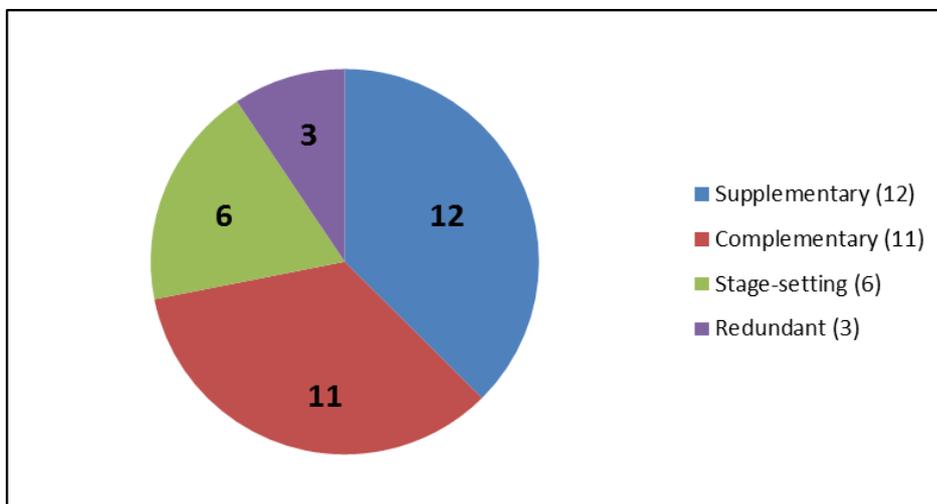


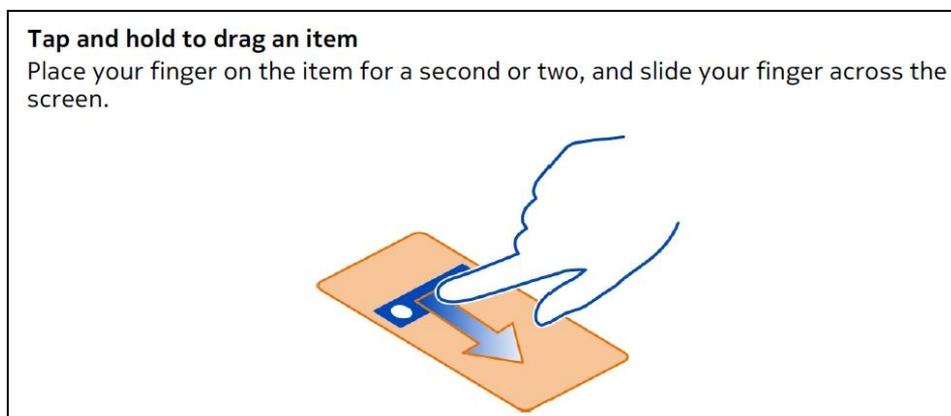
Figure 4: The relationships between text and images in *Lumia 800*'s user manual

The only relationship that does not appear in the user manual of *Lumia 800* is the juxtapositional relationship. When images and text are in a juxtapositional relationship, what the user imagines depends fully on the joint effect of text and images rather than on the interpretation of one or the other. The absence of the juxtapositional relationship is not surprising, because as stated in chapter 3.3, this relationship is most often used in advertising, for example. The purpose of user manuals is to instruct the users to use the device, not to bewilder the users with unexpected image-text combinations.

As the Figure 4 shows, the most common type of relationship is the supplementary relationship. This was actually quite an expected outcome, because the supplementary relationship tends to be the most commonly used way to integrate text and images: as pointed out in chapter 3.3, document designers tend to be a bit conservative in their image-text combinations as they most commonly use the supplementary relationship.

When inspecting the function of the images that form a supplementary relationship with the text, the following observation can be made: they are always either images that reinforce verbal description or images that express spatial relationships. Most often images in the supplementary relationship with the text are images that reinforce verbal description of action (10 out of 12 images). This kind of relationship can be seen in Example 11:

**Example 11: Supplementary relationship 1 (*Lumia 800*, page 20)**



Example 11 instructs the users how to drag an item on the touch screen. The text describes how to perform the action, while the image reinforces the instructions that the text gives. Also in Example 4 on page 44 there are two reinforcing images that form a supplementary relationship with the text. In all these types of relationships, images are the ones that supplement the text, whereas the text functions as the dominant mode. The frequent usage of reinforcing images in this relationship is in fact quite expected. As mentioned in chapter 3.2, in the supplementary relationship images often illustrate something that is hard to describe only with words. If the users have trouble imagining what is intended, supplementary images can help to clarify the content.

In addition to images that reinforce verbal description, images that express spatial relationships also form a supplementary relationship with the text in my material. Example 12 offers an example of this kind of relationship. Together the image and the text advise the user how to use the camera function of the phone:

**Example 12: Supplementary relationship 2 (*Lumia 800*, page 46; emphasis mine)**



- 1 To zoom in or out, select + or – in .
- 2 To focus, press and hold the camera key halfway down. The focus is locked when the white rectangle stops blinking.

**Tip:** When the focus is locked, you can keep the camera key pressed halfway down, and recompose the picture.

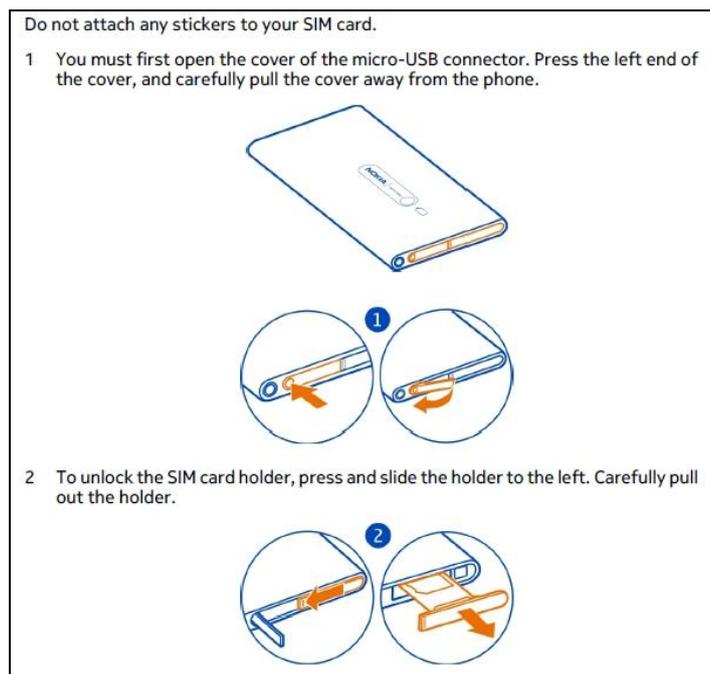
- 3 To take the picture, press the camera key fully down.  
Do not move the phone before the picture is saved and the final picture displayed.

The text functions as the main information source, whereas the image offers spatial clues about how users should keep the phone in their hand when taking a picture. In my opinion, the supplementary relationship between the text and the image is not a very successful choice here. The image does not add a lot of extra value to the information that the text already gives. Because the camera key is

positioned on the lower right side of the phone, the way in which the phone is held in the image is quite natural. So instead of depicting how to hold the phone, it would be more useful to show where the camera key, the white rectangle or other important parts of the phone, such as the camera lens and camera flash, are located, for example. Now the users must return to the very first image of the manual if they wish to know where these essential parts are located (Example 1 on page 41). At least there should be a reference that guides users to look at the first image of the manual if they are uncertain where the camera key is located, for instance. So the change that I would make is to replace the supplementary relationship with a complementary one by accompanying the text with an image that helps the users to locate all the crucial buttons and parts of the phone that are needed when taking a picture. That could be done by numbering all the parts to which the text refers and using those numbers in the image to make the location easier.

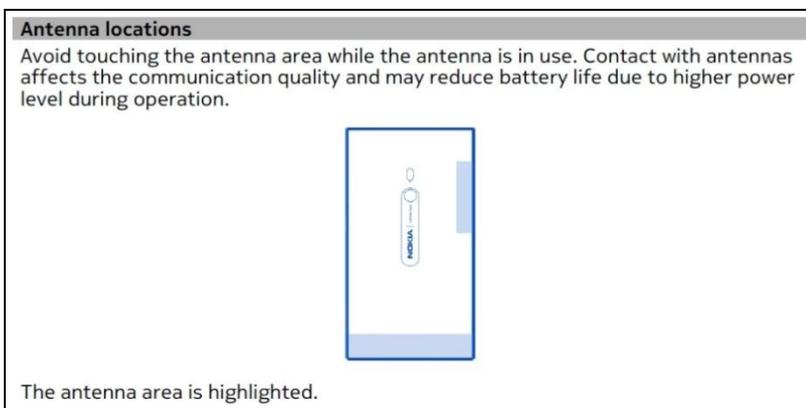
In fact, the complementary relationship is the second most common way in which text and images are integrated in the user manual (see Figure 4 on page 57). An example of a complementary relationship can be seen below in Example 13:

**Example 13: Complementary relationship 1 (*Lumia 800*, page 8)**



In Example 13 both the text and the images offer information that the other one does not provide. As mentioned in chapter 3.3, when words and images are in a complementary relationship, they can provide complete information about the action to take: the images give the user spatial cues about where to press or pull, while the text offers exact information about what to do and when to do it, which is exactly the case in Example 13. Another kind of example on the complementary relationship can be seen in Example 14 where the antenna areas of the phone are highlighted. Text and images complement each other and they could not function alone:

**Example 14: Complementary relationship 2 (*Lumia 800*, page 12)**



What is interesting to note here is that images that form a complementary relationship with the text are always the same kind of images: they all express spatial relationships. This once again shows how superior images are in expressing concrete objects, and on the other hand, how superior text is to images when it comes to describing things and actions. The text would not work without the image and vice versa.

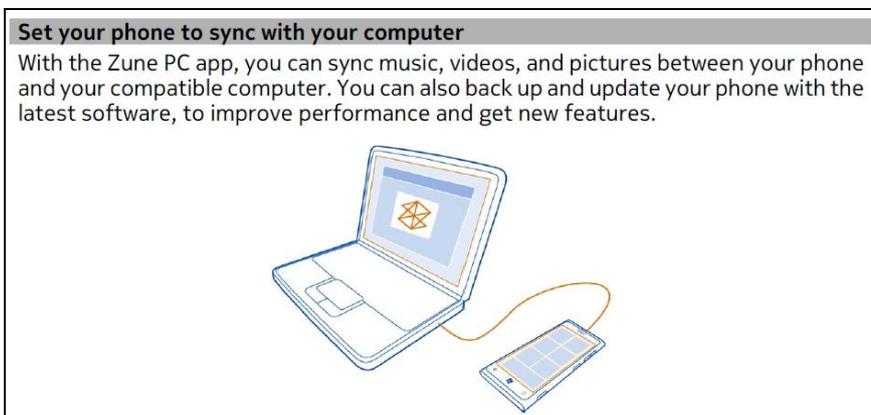
The third most common way in which text and images interact in the user manual is through a stage-setting relationship. As mentioned in chapter 3.3, stage-setting images forecast the theme of the text and help the users to get a sense of the big picture before they begin the reading. These kinds of relationships can be seen in Examples 15 and 16 on the next page. The examples instruct

the users to copy contacts from their old phone to the new one and to synchronise their phone with a computer:

**Example 15: Stage-setting image 1 (*Lumia 800*, page 15)**



**Example 16: Stage-setting image 2 (*Lumia 800*, page 18)**



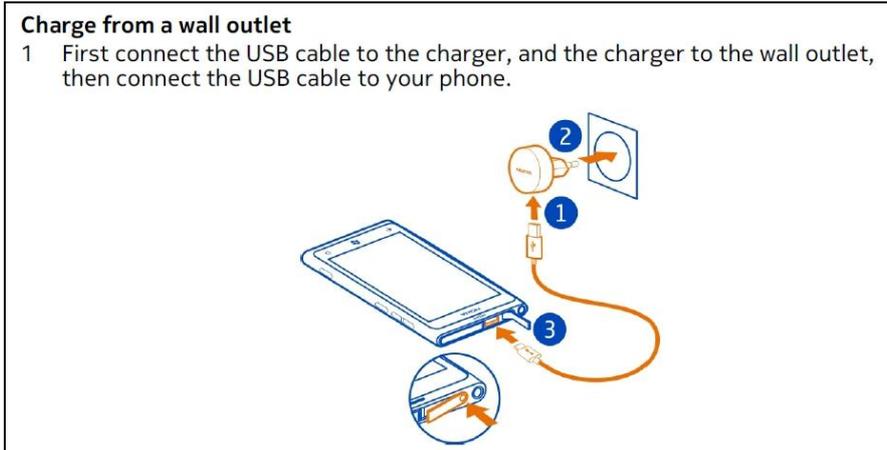
In both examples the image “sets the stage” for the upcoming text that gives more detailed information on the function. All stage-setting images in the user manual appear at the beginning of new chapters, which is actually one purpose to which Schriver (1997, 425) recommends stage-setting images. She states that “[i]t is common to conjoin the title of the chapter with an evocative illustration in the chapter’s opening spread.”

However, although images that form a stage-setting relationship with the text always appear in the same environment in the manual, it is complicated to distinguish these images from supplementary images. At first I could not decide whether text and images in Examples 15 and 16 interact through a supplementary or stage-setting relationship. That is to say, does the image add something to the text and support it, or does it merely depict what the text is about? I had this problem with some other image-text pairs as well, but upon closer examination, I was able to make the categorisation: if the image clearly does not add anything vital, such as spatial hints, to the text, I categorised it as a stage-setting image.

All of the images that form a stage-setting relationship with the text are orienting images. This is actually quite logical, because now that I have inspected the relationships that the images form with the text, I have come to the conclusion that the images in the manual that I categorised as orienting images are actually similar to images that Schriver (1997) calls stage-setting images. In Schriver's (1997, 425) words, stage-setting images provide a "visual anchor", which is exactly what orienting images do: they attach users' attention by providing an image that immediately tells what the text is about.

Furthermore, as pointed out in chapter 3.3, a stage-setting relationship can do more than just provide a visual anchor: sometimes the purpose of this relationship is to shape the users' attitude about the content in some particular way. This is the case with Example 15: the image gives a feeling that copying contacts is easy and effortless. The contacts practically fly from the old phone to the *Lumia 800*.

The final manner in which images and text interact in the user manual is through redundancy. Redundancy means that similar ideas are presented in alternative representations, in this case visually and verbally. In the user manual, the image and the text give exactly the same information only three times. In all these cases, the image and the text instruct the user to charge the phone. The image in Example 17 on the next page, as well as all the other images that form a redundant relationship with the text, give the users spatial clues to help the users to perform the task:

**Example 17: Redundant relationship (*Lumia 800*, page 10)**

In Example 17 the users are given instructions on how to charge the phone. Each phase that is described verbally is also expressed visually. As mentioned in chapter 3.3, redundancy is often used when it is hard for the user to fully understand a concept. However, as *Lumia 800* is aimed at people who are familiar with new technology, it is unlikely that charging the phone is challenging for them. Thus, the image would have been alone sufficient to describe the procedure and give spatial hints to the users.

To conclude, in the user manual of *Lumia 800* there are four ways in which text and images are integrated: supplementary, complementary, stage-setting and redundant. Images are most often used to supplement and complement the text: supplementary images reinforce textual description and give spatial clues, whereas complementary images are always used to express spatial relationships. Images that form a stage-setting relationship with the text are used at the beginning of new chapters to attract the attention of the users and to prepare the users to upcoming textual instructions. The least used integration method in the user manual is redundancy, which is used in connection with images that express spatial relationships. Redundancy is used only a few times, which is quite logical, considering that the main user group are people who are interested in new technology: as mentioned in chapter 3.3, redundancy can be a nuisance if the document designer shows the users something with which they are already familiar. Consequently, the excessive use of

redundancy can irritate the users and make the users think that the document designer underestimates them.

### 5.2.2 Gemini

As mentioned in chapter 3.3, the supplementary relationship tends to be the most commonly used integration method of text and image in user manuals. In this kind of relationship, images are most often the ones that support the text, that is to say, supplement the text. *Gemini*'s user manual is no exception: as Figure 5 shows, the supplementary relationship is clearly the most often used relationship, and it is always the images that supplement the text and not vice versa:

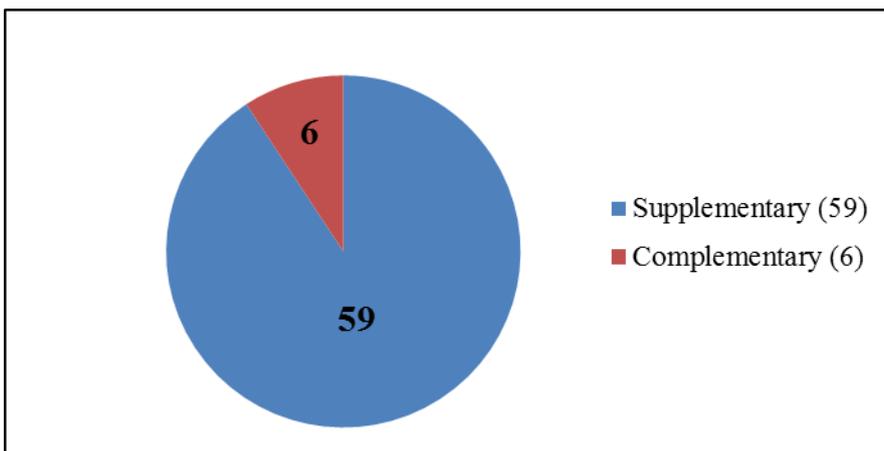


Figure 5: The relationships between text and images in *Gemini*'s user manual

The high number of supplementary images correlates with the frequent usage of verifying images: almost all supplementary images in the user manual are images that verify screen states. An example of the supplementary relationship between the text and the image can be seen in Example 18 on the next page:

**Example 18: Supplementary relationship (Gemini)**

1. Go to **Projects > Delete Unaccepted Project Assignment**.  
The Delete a Purchase Order window opens.

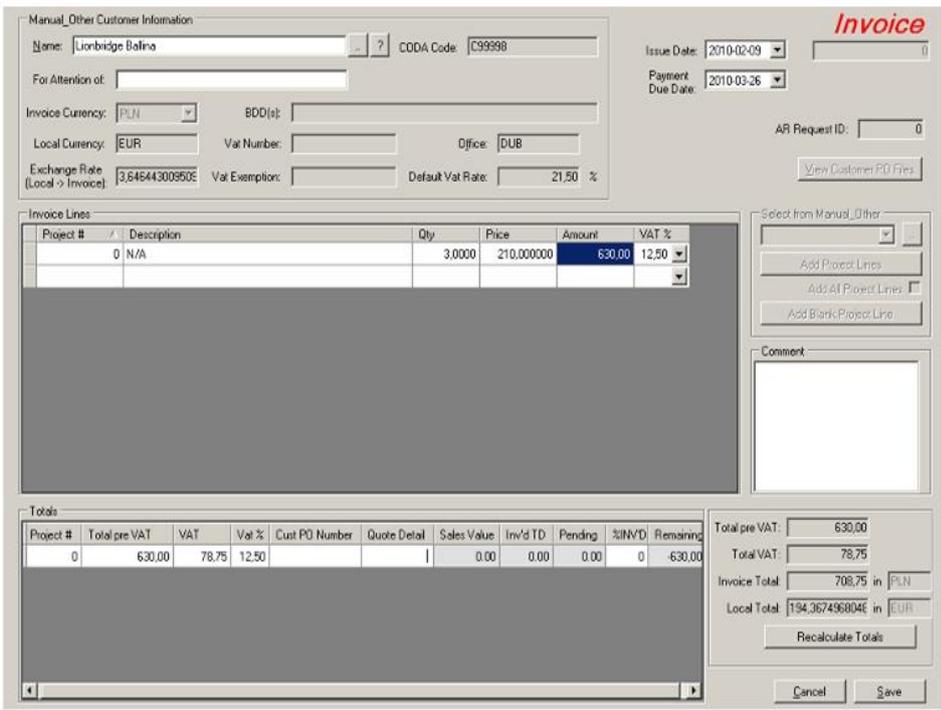


2. Select the PA with the **PO/PA Picker** (the .. button).  
The Select a Purchase Order or Project Assignment window opens.

The instruction advises the users how to delete project assignments. After the first step, the text informs the users that the “Delete a Purchase Order” window opens and the image supports the textual information. As mentioned in chapter 5.1.2, these kinds of images confirm the users’ actions by displaying the windows that the users should see on their screens. In cases such as Example 18, the supplementary image is a reasonable choice to accompany the text, because the image works well for verifying the users’ actions and there is no need for other kind of support. However, there are several instances in the manual where the supplementing images do support the users’ learning process as effectively as possible, and the reason for this is the complexity of the window. An example of this kind of situation can be seen on the next page:

**Example 19: Insufficient use of the supplementary relationship (Gemini)**

The AR Creation window opens.



The screenshot shows a software window titled 'Invoice' with the following sections:

- Manual\_Other Customer Information:** Name: Lionbridge Balina, CODA Code: C99998, Issue Date: 2010-02-09, Payment Due Date: 2010-03-26, AR Request ID: 0.
- Invoice Currency:** PLN, BDD(s):, Local Currency: EUR, Office: DUB, Exchange Rate (Local > Invoice): 3,64644300950E, Default Vat Rate: 21,50 %.
- Invoice Lines:** A table with columns: Project #, Description, Qty, Price, Amount, VAT %. One row is visible: Project # 0, Description N/A, Qty 3,0000, Price 210,000000, Amount 630,00, VAT % 12,50.
- Totals:** A table with columns: Project #, Total pre-VAT, VAT, Vat %, Cust PO Number, Quote Detail, Sales Value, Inv'd TD, Pending, %INVD, Remaining. One row is visible: Project # 0, Total pre-VAT 630,00, VAT 78,75, Vat % 12,50, Sales Value 0,00, Inv'd TD 0,00, Pending 0,00, %INVD 0, Remaining 630,00.
- Summary:** Total pre VAT: 630,00, Total VAT: 78,75, Invoice Total: 708,75 in PLN, Local Total: 194,367496804€ in EUR.

This window is similar to the Gemini invoice creation window, but the selected Gemini project field is grayed out and you cannot add Project Lines to the invoice.

2. Enter the name of the customer, or click the **Customer Picker** (the .. button) and select the customer.
3. Add lines to the invoice manually.  
Enter a description, quantity and price, and choose the correct VAT rate for the invoice line. Add as many lines as required.
4. To view the invoice line totals, click **Recalculate totals**.

The first supplementing screen capture (Example 18) is considerably smaller and more simple than the second example (Example 19). These differences in size and complexity define the sufficiency of the supplementary relationship: in the second example, the image is crowded with different kinds of sections and buttons that the text describes. Thus, the supplementary relationship between the text and the screen capture does not serve the need which users most likely have: the need to locate different objects effortlessly. The supplementary image functions well for verifying the users' actions, but it does not make it any easier for the users to locate objects. In these kinds of situations, an image that helps users to locate objects would be a good choice to complement the text.

As mentioned in chapter 5.1.2, locating screen captures are especially important when windows are crowded with elements. By adding step numbers in the screen capture to indicate the location of different objects can make the location process faster and reduce errors, for example.

This kind of screen capture is actually used once in the manual. This example occurs in connection with the introduction of the Gemini Online Help user interface:

**Example 20: Efficient utilising of the complementary relationship (*Gemini*)**

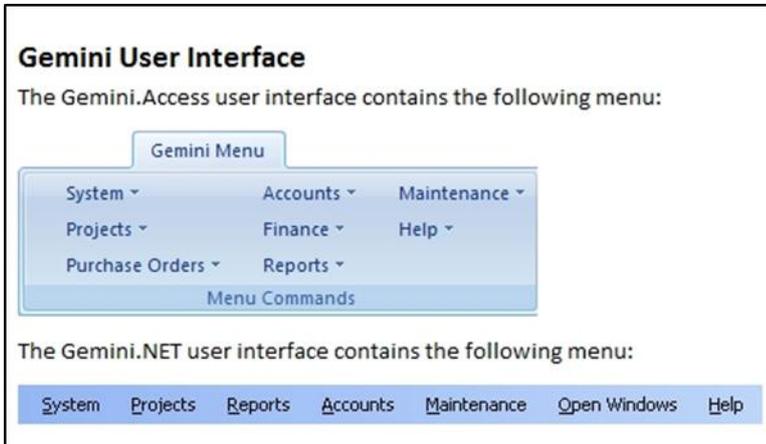


The numbers help the user to locate the relevant buttons and menus and help the users to focus their attention faster on the relevant parts of the window. The same kind of relationship would make Example 19 much more profitable for the users of *Gemini*. If the supplementary image that verifies screen states was replaced with the complementary image that expresses location, the image would not merely support the text but would provide information that the text does not provide: information about the spatial relationships of the objects on the screen.

As Figure 5 on page 65 shows, the complementary relationship is utilised six times in the user manual of *Gemini*. In three of six cases, text forms a complementary relationship with the image that expresses location, as in Example 20. In addition, text forms a complementary relationship with

the images that help the users to build a mental model. Example 21, which has already been introduced in the analysis chapter 5.1.2, is a good example on this kind of relationship:

**Example 21: Complementary relationship (*Gemini*)**



In Example 21, the texts and the images provide different visual and verbal content: the text describes and the image depicts the idea. As mentioned in chapter 3.3, when text and images are in a complementary relationship, they clarify and strengthen the users' understanding of the main idea. In Example 21, text and images together help the users to acquaint themselves with the structure of the user interface. The images would not work without the texts and the texts would not work without the images.

To conclude, the use of the supplementary relationship is clearly the most dominant way to integrate text and images in the user manual of *Gemini*. In all instances, images are the ones that supplement the text and not vice versa. The only other kind of way in which text and images are integrated is through complementary relationship. However, the use of the complementary relationship is marginal when compared with the use of the supplementary relationship. Consequently, in the user manual of *Gemini* images take the supportive role while text carries the more vital information.

I do not claim that using the supplementary relationship in user manuals is always unhelpful. However, I would claim that in the case of this specific user manual the frequent usage of the

supplementary images do not always give the best possible support for the users. The user manual supports the users in a one-sided manner and does not fulfil all the users' needs. As mentioned in chapter 2.1, the aim of document design is to produce texts that help people to achieve their goals. Consequently, when creating instructions, document designers must keep the users' tasks in mind and make textual and visual choices that best help the users to complete those tasks. Screen captures are no exception on this issue: Gellevij and van der Meij (2004, 235) point out that if the goal of documentation is to support the tasks that the users need to complete, screen captures should also be used only when users benefit from their presence. In practice, this means that document designers need to carefully analyse the users' task to be able to decide what kinds of screen captures should be used and how the screen captures should be integrated with the text.

Horton (1993, 146) offers an interesting view on this issue. He argues that users can get frustrated with screen captures because they do not add anything new to what users already see on the screen. Horton aptly points out that "What's it look like?" is not always the questions to which the users seek answers for. One possible question that Horton mentions is: "Where am I?"

In fact, Horton (1993) highlights the same kind of idea that van der Meij and Gellevij (1998) bring forward in their article: the users' needs should guide the use of the screen captures. Whether the users' question is "What's it look like?" or "Where am I?", the screen captures need to help the users to get answers to their questions. However, quite often in the user manual of *Gemini* when users most probably need help in locating objects on the screen, the images merely support the verification of the screen states. In my opinion, this is just the kind of situation where the supporting image should be replaced by an image that complements the text by expressing location. By using the complementary relationship to integrate text and images, the text describes **what** the users need to do and the image shows **where** the users need to focus their attention.

It is actually rather surprising that the user manual of *Gemini* does not help the users in location tasks, whereas the user manual of *Lumia 800* does it several times. Although *Lumia 800* and *Gemini* represent different product types, hardware and software, I would argue that facilitating

the users' locating process is equally important in both manuals. After all, by helping the users to locate objects more effectively, the manual reduces the cognitive load of the users. As mentioned in chapter 5.1.2, this in turn speeds up the tasks that the users need to complete and reduces the errors that they make.

All in all, there are differences in the integration of text and images between the two studied user manuals. In the user manual of *Lumia 800*, different kinds of relationships between text and images are used much more diversely, as the manual takes advantage of the supplementary, complementary, stage-setting as well as redundant relationships. In the user manual of *Gemini* only the first two relationships are used.

Naturally, it is impossible to directly compare these totally different types of user manuals. It is obvious that because software and hardware user manuals differ in the kinds of images that are used, images are also integrated differently with the text. To begin with, the fact that almost all images in the user manual of *Gemini* are images that verify screen states is obviously the reason why the supplementary relationship is so dominant. The supplementary relationship is also the most frequent way to integrate text and images in *Lumia 800*'s user manual: the supplementary relationship is most often used when the text describes an action which needs to be reinforced with the help of an image. So in both user manuals, the supplementary relationship is utilised when actions need to be confirmed: in the user manual of *Gemini*, the supplementary screen captures verify that the users have done the right thing, whereas in the user manual of *Lumia 800*, the image confirms that the users understand the actions that the text describes.

A practical conclusion that can be drawn is that both in hardware and software documentation, images seem to work well in a supplementary relationship with the text when it comes to expressing action. Text functions as the dominant mode, while images make sure that the users understand the action that they need to complete. This, of course, means that in order to know when images are needed to supplement text, document designers need to have information on the knowledge and skills of the users.

The complementary relationship has also been taken advantage of in both user manuals. In both manuals the complementary relationship is used to express spatial relationship and location. On the basis of the analysis of the two user manuals, I would claim that because images are good at expressing location while text is good at describing difficult series of actions, the complementary relationship in locating tasks offers an effective way to integrate text and image in both software and hardware manuals.

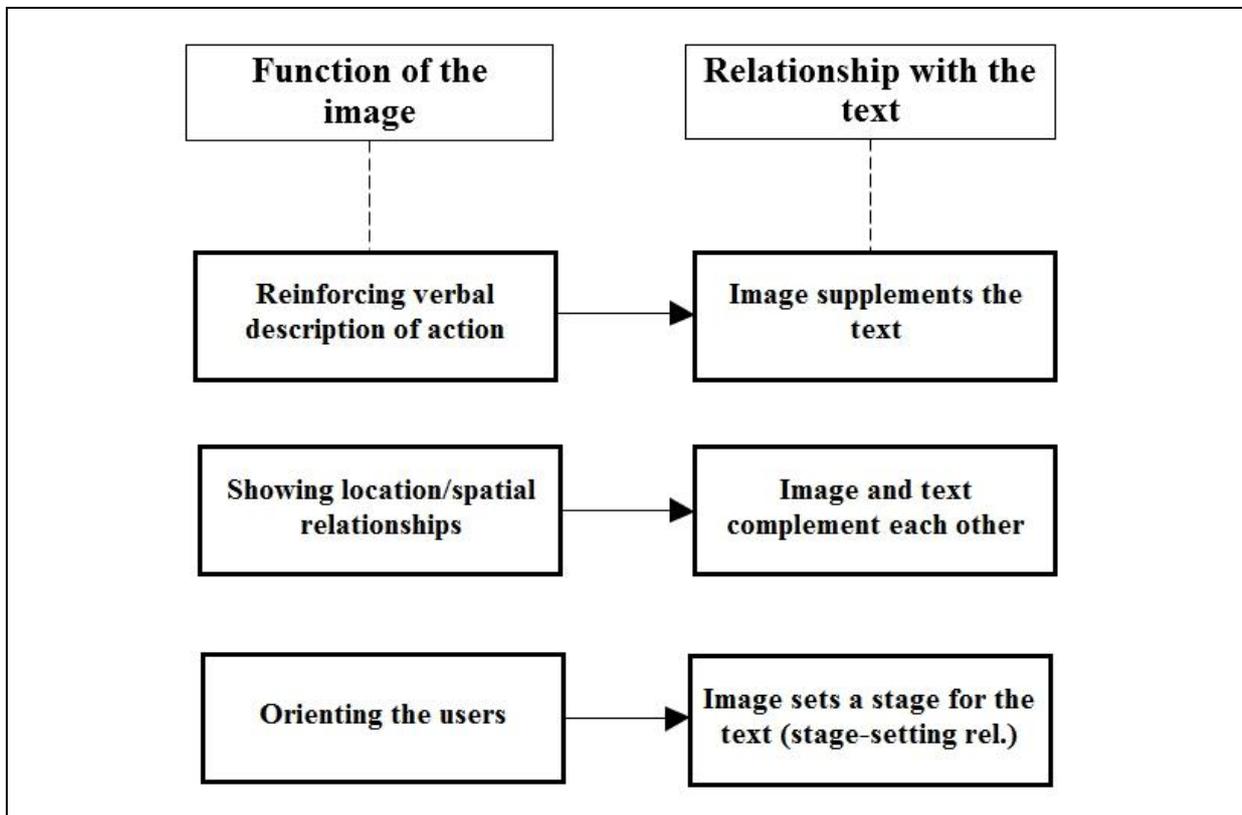
In addition, the user manual of *Gemini* complements the textual description with images that help the users to build a mental model of the programme. The complementary text and image combination is used to explain the basic structure of the programme. Based on the analysis, this kind of use of the complementary relationship between text and images seems to be useful in software manuals. As mentioned in chapter 5.1.2, it is important to help users to develop an understanding of the basic structure of the programme, because without knowledge about the design of the screens, the users would constantly be overwhelmed by the information they see on the screen.

In the user manual of *Lumia 800*, the redundant relationship is also used a few times to integrate text and images. All of the images that form the redundant relationship with the text are images that express spatial relationships. However, I would not generalise that redundant relationship works well for integrating text with images that give spatial clues. As mentioned in chapter 3.3, the redundant relationship is often used when something particularly complex needs to be expressed. Thus, the reason why the redundant relationship is so seldom used is most probably the fact that neither of the studied manuals is aimed at beginners. Thus, whether to use the redundant relationship or not depends on how much knowledge the users have, in other words, whether the users are experts, novices or something in between.

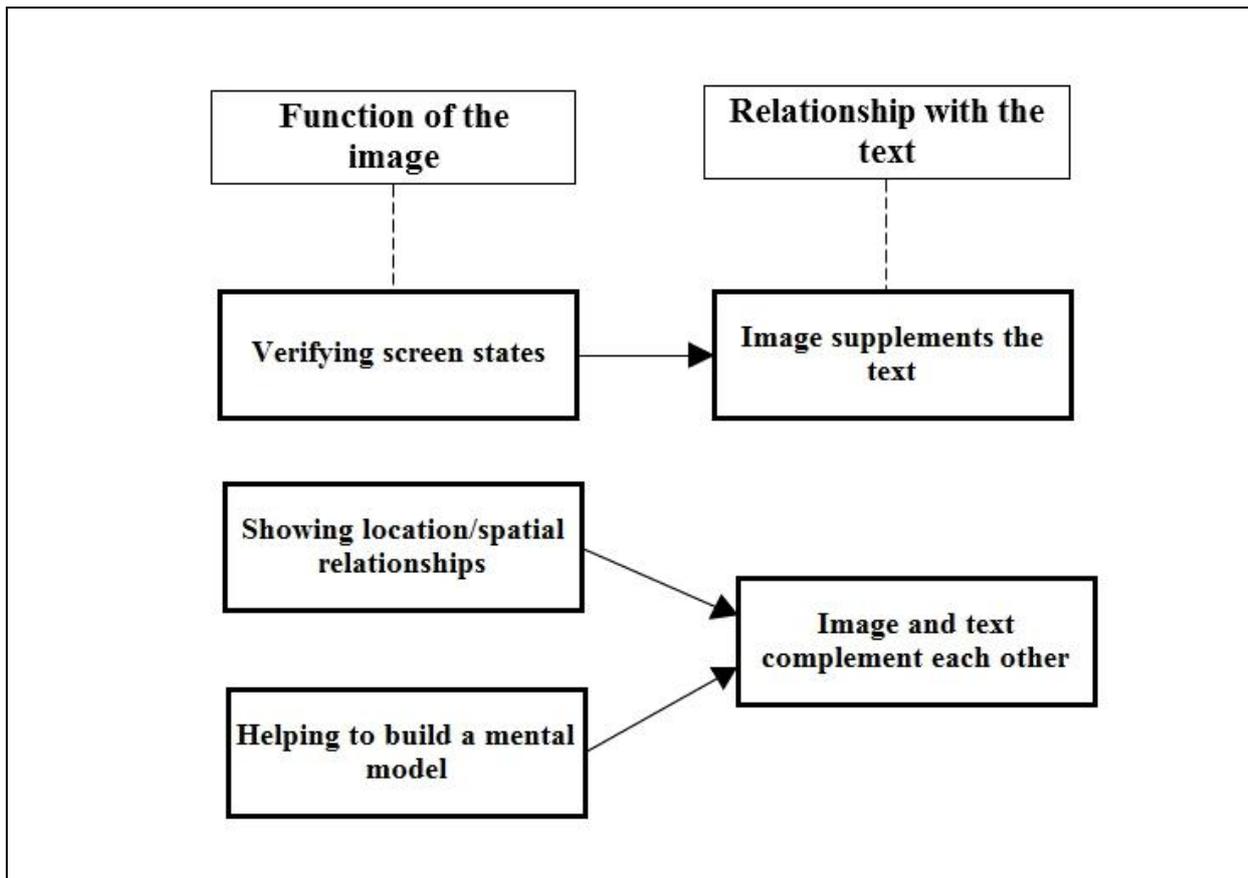
*Lumia 800* also quite often takes advantage of the stage-setting relationship between text and images. These images catch the users' eye by providing a visual anchor that immediately tells the user what the chapter is about. However, these kinds of images presumably work best in hardware

documentation: as the user manuals of hardware products, such as of *Lumia 800*, describe concrete objects, it is easy to use images to visualise the theme of the chapter, because images are good at expressing concrete objects. However, it would not be so easy to use an image to set a stage for the upcoming content in a software manual, such as *Gemini*, because images do not function so well in describing abstract ideas.

Figures 6 and 7 summarise the results of the analysis. The left-hand side of the column shows the purposes to which images are used in the manuals, whereas the right-hand side of the column indicates how the images can effectively be integrated with the text:



**Figure 6: Guidelines for the integration of text and images in hardware documentation**



**Figure 7: Guidelines for the integration of text and images in software documentation**

These results concerning the two studied user manuals can offer suggestive guidelines for document designers when they create documents. The guidelines that Figures 6 and 7 depict can easily be applied to user manuals of other similar kinds of products as well. The recommendations that were made on the basis of *Lumia 800*'s user manual could also be used to guide the design process of other similar types of hardware products, such as other phone models, tablet computers, GPS systems or portable devices in general. *Gemini*'s user manual, on the other hand, can offer relevant guidelines for designing user manuals that describe software with similar kind of user interface, for example.

## 6 Conclusions

In this thesis I have studied the interplay of text and images in two user manuals, *Lumia 800* and *Gemini*. I have analysed the purposes for which images are used and how these images are integrated with the text. I wanted to find out how document designers can take advantage of images in documents.

To begin with, the analysis shows that the two studied user manuals differ in their use of images: images in *Lumia 800*'s user manual focus on expressing spatial relationships, reinforcing verbal descriptions and orienting the users, whereas images in the user manual of *Gemini* focus on clarifying abstract content by describing the structure of the software and by verifying that the users keep up with the instructions. In addition, images in *Gemini*'s user manual also help the users to locate different objects on the window. Thus, whether “a picture is worth a thousand words” is highly dependent on the idea that needs to be expressed.

However, as the theoretical background of this study shows, it is not enough for document designers to understand how words and images work separately, but they also need to know how they act interdependently. The user manual of *Lumia 800* takes advantage of the supplementary, complementary, stage-setting as well as redundant relationship. On the contrary, in the user manual of *Gemini* only the first two relationships are used.

These differences in integration methods clearly reflect the fact that different document types use images for different purposes. As the user manual of *Gemini* describes abstract software, it is natural that text, which works better for describing abstract concepts, takes the more dominant role in the manual, whereas images are used to support the text. On the contrary, the user manual of *Lumia 800* gives instructions about a concrete object, a phone. Consequently, as images are often better than text at expressing concrete objects, the manual does not merely use images to support the text but to add something to it as well.

By inspecting the connections between the functions of images and the relationships they form with the text, interesting conclusions can be drawn: in both user manuals, images work well in

a supplementary relationship with the text when it comes to describing action, whereas a complementary relationship works best when users need to locate something. In the user guide of *Gemini*, the complementary relationship is also used to help the users to build a mental model of the programme. The user guide of *Lumia 800* also takes advantage of the stage-setting and redundant relationship: images that form a stage-setting relationship with the text are good at orienting the users, whereas redundancy between text and images is useful when document designers want to convey a complex idea.

These results provide answers to the questions raised at the outset: I have been able to form suggestive recommendations for the effective integration of text and images. I believe that these results are useful to the document designers of Lionbridge, as well as to document designers in general. As already mentioned, although achievements in the research on multimodality have been quite remarkable, the interplay between different modes have not received much attention. Previous studies have mainly focused on the function of individual modes. So this study responds to the growing need for knowledge on the integration of text and images in technical documentation.

The method used in this study proved to be rather challenging, because sometimes I had problems in distinguishing between two different categories. However, as the results show, the method was sufficient to achieve the aim of this study. Especially the decision to use different models for inspecting the functions of images in the two manuals proved to be a good choice, because screen captures in fact have a function that does not apply to images in general: they are used to verify screen states.

However, as the material of this study is not very extensive, it is presumable that there are many different kinds of text-image relationships that did not appear in the studied user manuals. The two user manuals proved to be a suitable amount of material for this study, but naturally, the material is insufficient to make any generalisations on the effective integration of text and images. So more studies on the integration of different modes in multimodal documents are clearly needed. An interesting topic for further research could be, for example, to conduct a usability test to find out

how users react to the text-image guidelines discovered in this study. In addition, as this study highlights the effect of the product type on the integration of text and images, it would be useful to pay attention to the other affecting factors as well. For example, we could study how much the company's conventions, cultural characteristics or document designers' attitudes affect the integration of text and images.

All in all, as the growing importance of visuals in our culture challenges document designers to inspect the relationships between text and images more closely, further studies on the integration of text and images are clearly needed. By focusing on text-image integration, this study has scratched the surface of a wide research field of multimodality in document design, and although several interesting results were achieved, yet several aspects remain to be studied further. However, I hope that this study offers practical guidelines for document designers and encourages them to view the integration of text and images in a new light.

## Works Cited

### Primary Material

*Lumia 800* user manual. [Internet] Available from [http://nds1.nokia.com/phones/files/guides/Nokia\\_800\\_UG\\_en.pdf](http://nds1.nokia.com/phones/files/guides/Nokia_800_UG_en.pdf). [Accessed 10 November 2012]

*Gemini* user manual. Lionbridge internal material.

### Secondary Sources

Aarhus School of Business. 2009. Three LBC researchers win prestigious distinguished awards. [Internet] Available from [http://www.asb.dk/en/aboutus/departments/departmentofbusinesscommunication/news/article/artike/l/three\\_lbc\\_researchers\\_win\\_prestigious\\_distinguished\\_awards-2/](http://www.asb.dk/en/aboutus/departments/departmentofbusinesscommunication/news/article/artike/l/three_lbc_researchers_win_prestigious_distinguished_awards-2/). [Accessed 25 February 2012]

Bezemer, Jeff and Gunther Kress. 2008. "Writing in Multimodal Texts: A Social Semiotic Account of Designs for Learning." *Written Communication* 25: 166–195.

Björkvall, Anders. 2009. *Multimodal Analys i Praktiken*. Stockholm: Hallgren & Fallgren.

Booher, Harold R. 1975. "Relative Comprehensibility of Pictorial Information and Printed Words in Proceduralized Instructions." *Human Factors* 17, 3: 266–277.

Brumberger, Eva R. 2007. "Making the Strange Familiar. A Pedagogical Exploration of Visual Thinking." *Journal of Business and Technical Communication* 21, 4: 376–401.

Donnel, Jeffrey. 2005. "Illustration and Language in Technical Communication." *Journal of Technical Writing and Communication* 35, 3: 239–271.

Doumont, Jean-Luc. 2002. "Verbal Versus Visual: A Word Is Worth a Thousand Pictures, Too." *Technical Communication* 49, 2: 219–224.

Fortune, Ronald. 2002. "Image, Word, and Future Text. Visual and Verbal Thinking in Writing Instruction." In *Working with Words and Images. New Steps in an Old Dance*, ed. Nancy Allen, 97–115. Westport, Connecticut/London: Ablex Publishing.

Ganier, Franck. 2004. "Factors Affecting the Processing of Procedural Instructions: Implications for Document Design." *IEEE Transactions on Professional Communication* 47, 1: 15–26.

Ganier, Franck. 2009. "Observational Data on Practical Experience and Conditions of Use of Written Instructions." *Journal of Technical Writing and Communication* 39, 4: 401–415.

Gellevij, Mark and Hans van der Meij. 2004. "Empirical Proof for Presenting Screen Captures in Software Documentation." *Technical Communication* 51, 2: 224–238.

Gribbons, William M. and Arthur G. Elser. 1998. "Visualizing Information: An Overview of This Special Issue." *Technical Communication* 45, 4: 467–472.

Gurak, Laura J. 2003/1992. "Toward Consistency in Visual Information: Standardized Icons Based on Task." *Technical Communication* 50, 4: 492–496.

Harrison, Claire. 2003. "Visual Social Semiotics: Understanding How Still Images Make Meaning." *Technical communication* 50, 1: 46–60.

Horton, William. 1992. "Pictures Please – Presenting Information Visually." In *Techniques for Technical Communicators*, ed. Carol M. Barnum and Saul Carliner, 188–219. New York: Macmillan.

Horton, William. 1993. "Visual Literacy. Going beyond Words in Technical Communications." *Technical Communication* 40, 1: 146–148.

Johnson-Sheehan, Richard. 2002. "Being Visual, Visual Beings." In *Working with Words and Images. New Steps in an Old Dance*, ed. Nancy Allen, 73–96. Westport, Connecticut/London: Ablex Publishing.

Klink, Mitch. 2000. Technical Writing in the 21st Century. [Internet] The Willamette Galley. A Quarterly Newsletter 3, 5. Available from [http://www.stcwvc.org/galley/Nov00\\_newsletter/Writing-21Century.htm](http://www.stcwvc.org/galley/Nov00_newsletter/Writing-21Century.htm). [Accessed 28 September 2011]

Kostelnick, Charles. 1994. "From Pen to Print: The New Visual Landscape of Professional Communication." *Journal of Business and Communication* 8: 91–117.

Kress, Gunther. 1998. "Visual and Verbal Modes of Representation in Electronically Mediated Communication: The Potentials of New Forms of Text." In *Page to Screen. Taking Literacy into the Electronic Era*, ed. Ilana Snyder, 53–79. London/New York: Routledge.

Kress, Gunther. 2000. "Multimodality." In *Multiliteracies. Literacy Learning and the Design of Social Futures*, ed. Bill Cope and Mary Kalantzis, 182–202. London & New York: Routledge Taylor & Francis Group.

Kress, Gunther. 2003. *Literacy in the New Media Age*. London: Taylor & Francis.

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

Lehtonen, Mikko. 2002. "Surmaako kuva sanan? Multimodaalisuuden haasteet tekstintutkimukselle." In *Kieli yhteiskunnassa – yhteiskunta kielessä. AfinLAN vuosikirja 2002 / n:o 60*, ed. Anna Mauranen & Liisa Tiittula, 45–60. Jyväskylä: Suomen soveltavan kielitieteen yhdistys AfinLA.

Lewis, Elaine. 1988. "Design Principles for Pictorial Information." In *Effective Documentation. What We Have Learned from Research*, ed. Stephen Doheny-Farina, 235–253. Cambridge: The MIT Press.

Lim, Fei V. 2004. "Developing an Integrative Multi-Semiotic Model." In *Multimodal Discourse Analysis*, ed. Kay O'Halloran, 220–244. London: GBR: Continuum International Publishing.

Loorbach, Nicole, Karreman, Joyce and Michaël Steehouder. 2007. *The Effects of Adding Motivational Elements to User Instructions*. Professional Communication Conference.

Lu, Xiaonan, Kataria, Saurabh, Brouwer, William J., Wang, James Z., Mitra, Prasenjit and C. Lee Giles. 2009. Automated Analysis of Images in Documents for Intelligent Document Search. [Internet] Available from <http://infolab.stanford.edu/~wangz/project/imsearch/DOCIMG/IJDAR09/lu.pdf>. [Accessed 8 January 2012]

Maier, Carmen D., Kampf, Constance and Peter Kastberg. 2007. "Multimodal Analysis: An Integrative Approach for Scientific Visualizing on the Web." *Journal of Technical Writing and Communication* 37, 4: 453–478.

Nielsen, Jakob. 2000a. Is Navigation Useful? [Internet] Available from <http://www.useit.com/alertbox/20000109.html>. [Accessed 29 April 2012]

Nielsen, Jakob. 2000b. Eyetracking Study on Web Readers. [Internet] Available from <http://www.useit.com/alertbox/20000514.html>. [Accessed 29 April 2012]

Pfeiffer, William S. 2000. *Technical Writing. A Practical Approach*. 4<sup>th</sup> ed. New Jersey: Prentice Hall.

Portewig, Tiffany. 2004. "Making Sense of the Visual in Technical Communication: A Visual Literacy Approach to Pedagogy." *Journal of Technical Writing and Communication* 34, 1&2: 31–42.

Portewig, Tiffany. 2008. "The Role of Rhetorical Invention for Visuals: A Qualitative Study of Technical Communicators in the Workplace." *Technical Communication* 55, 4: 333–342.

Pulkkinen, Matti. 20.12.2011. E-mail conversation concerning the two studied user manuals.

Qiuye, Wang. 2000. "A Cross-cultural Comparison of the Use of Graphics in Scientific and Technical Communication." *Technical Communication* 47, 4: 553–558.

Schraver, Karen A. 1997. *Dynamics in Document Design*. New York: John Wiley & Sons, Inc.

STVY. Suomen teknisen viestinnän yhdistys. [Internet] Homepage. Available from <http://www.stvy.fi/node/8>. [Accessed 26 March 2012]

Van der Meij, Hans and Mark Gellevij. 1998. "Screen Captures in Software Documentation." *Technical Communication* 45, 4: 529–543.

Van der Meij, Hans and Mark Gellevij. 2002. "Effects of Pictures, Age, and Experience on Learning to Use a Computer Program." *Technical Communication* 49, 3: 330–339.

Ventola, Eija, Cassily, Charles and Martin Kaltenbacher. 2004. *Perspectives on Multimodality*. Amsterdam/Philadelphia: John Benjamins Publishing Company.

Ware, Collin. 2004. *Information Visualization. Perception for Design*. 2<sup>nd</sup> ed. San Francisco: Morgan Kaufmann publications.

Willerton, Russell. 2005. "Visual Metonymy and Synecdoche: Rhetoric for Stage-setting Images." *Journal of Technical Writing and Communication* 35, 1: 3–31.

Williams, Thomas R. and Deborah A. Harkus. 1998. "Editing Visual Media." *IEEE Transactions on Professional Communication* 41, 1: 33–46.