

ANNE KETOLA

# Word–Image Interaction in Technical Translation

Students translating an illustrated text



ANNE KETOLA

Word–Image Interaction  
in Technical Translation

Students translating an illustrated text



ACADEMIC DISSERTATION  
To be presented, with the permission of  
the Faculty Council of the Faculty of Communication Sciences  
of the University of Tampere, for public discussion  
in the auditorium Pinni B 1100, Kanslerinrinne 1, Tampere,  
on 24 April 2018, at 12 o'clock.

UNIVERSITY OF TAMPERE

ANNE KETOLA

Word–Image Interaction  
in Technical Translation

Students translating an illustrated text

*Acta Universitatis Tamperensis 2364*  
*Tampere University Press*  
*Tampere 2018*

## ACADEMIC DISSERTATION

University of Tampere  
Faculty of Communication Sciences  
Finland

The originality of this thesis has been checked using the Turnitin OriginalityCheck service in accordance with the quality management system of the University of Tampere.

Copyright ©2018 Tampere University Press and the author

Cover design by  
Mikko Reinikka

Acta Universitatis Tamperensis 2364  
ISBN 978-952-03-0705-9 (print)  
ISSN-L 1455-1616  
ISSN 1455-1616

Acta Electronica Universitatis Tamperensis 1870  
ISBN 978-952-03-0706-6 (pdf)  
ISSN 1456-954X  
<http://tampub.uta.fi>

Suomen Yliopistopaino Oy – Juvenes Print  
Tampere 2018



# ACKNOWLEDGEMENTS

The section you are about to read contains a thank-you directed towards ninety-eight individuals. I understand that this sounds ridiculous but I beg my reader to bear with me. I will start by thanking the people who have directly contributed to the outcome of this research project. I thank my supervisors, Dr. Riitta Oittinen and Dr. Tiina Tuominen, for the indispensable help I have received during this project. I am proud to notice that, during the past years, in addition to being colleagues, we have also become friends. This I value even more than the continuous support I received whenever I needed it. I am grateful to Prof. Kaisa Koskinen for the insightful comments that helped me significantly improve the last part of this dissertation. I want you to know that your support has meant a lot to me. I am thankful to Dr. Carol O'Sullivan and Prof. Hanna Risku who acted as the preliminary evaluators of this dissertation, and to the four editors and the eight reviewers who helped improve the articles that constitute the basis of this study. I wish to thank Dr. Hilkka Pekkanen for allowing me to visit her translation seminar for data collection purposes in 2014 and I thank the students who took part in the study. I also owe my thanks to the staff at the Eastern Finland office of the Geological Survey of Finland for their help in creating the source text that was used in this research, and to Dr. Richard Browner from the Western Australian School of Mines for kindly proofreading it.

Somewhat paradoxically, I believe that my PhD project has been sped up by taking on other, smaller research projects on the side, most of which I have conducted in collaboration with others. This has allowed me to work with various brilliant colleagues: in addition to Tiina and Riitta mentioned above, this list includes Eliisa, Sari, Maija, Nina, Melissa, Chiara, Roberto, George, Alejandro and Catalina. Working with you has challenged my own thinking and deepened my understanding of the discipline. I sincerely thank you for this.

I first heard of magnetic ore beneficiation – the subject matter of the source text used in my research – during a 10-hour car journey from Tampere to Lapland. My father, my brother and I were on our way to our cabin for Easter holidays in the spring of 2013. My father, something of an inventor himself, is fascinated by how things work, and will often engage in painstakingly detailed monologues about the

operating principles of things. During this particular trip, we received what felt like a two-hour-long description of magnetic and non-magnetic ore particles moving in slurry. Even though, at the time, I might not have appreciated the lecture (and, I suspect, I might have begged him to stop), this bit of information turned out to be quite valuable indeed. For this inspiration, Dad, I say thanks.

I also want to express my gratitude to the rest of my family. Mum, I suspect that the very first seed of my academic aspirations was sown in 2003 when you defended your PhD. I will never understand how you were able to carry out such a substantial piece of work while having a full-time job and three needy kids. When people compliment me on my drive and my determination, I cannot help thinking that, with a role model like you, I could not have turned out any other way. I thank my sister Hanna, for your warm support, and my brother Antti, for your refreshing noninterest towards my work. You two mean the world to me. I am grateful to, and grateful for, my awesome nephews, Eetu and Otto. One of my favourite things about the past four and a half years has been witnessing you turn from toddlers to wonderful little gentlemen. Being your godmother is one of my greatest prides in life. I thank my lovely mother-in-law Riitta for the countless baby-sitting favors, without which this project would have taken a lot longer. I am particularly grateful to my husband Raino, for being such a great friend and such an amazing father, and for tolerating my desire to keep up with the sometimes ridiculous pace of academic work. Finally, my sincerest thank-you goes to our beautiful son Roope, for keeping me grounded, and for reminding me, every single day, that there are countless things in life that are infinitely more important than research. I don't know what I've done to deserve someone like you in my life. You are precious, inside and out. Äiti rakastaa sinua.

I also wish to send a warm thank-you to those who have colored my non-academic life in the past four and a half years. During this time, I have had the pleasure of working and spending time with extraordinarily passionate and talented people. Roxana, Zoila, Rosamaría, Aya, Mohsen, Saad, Ye Yint – it has been an honor to translate your amazing work. Our projects have allowed me to fully understand and appreciate the *impact* of translation, both on a personal as well as a social level. I am beyond exited to be starting the new chapter on our quest. I am also grateful to the group of budding artists I have had the pleasure of creating and studying visual art with in the past few years. Our conversations have helped me construct a more thorough understanding of *how images mean*; what images are, and what they are not. I thank Pelle for filling my heart with images. I send a special thank-you to Hennu and Heidi, for showing me what it looks like when a person

follows her heart and lives without fear. I also send a glittery thank-you to my bizarrely beautiful burlesque family, for showing me that few are the problems that cannot be fixed with a bit of bright lipstick and a bird costume. Last but by no means the least, I thank my lovely Mu-Tan Clan for twenty years of friendship – and the years to come.

Tampere, March 2018

Anne Ketola



# TIIVISTELMÄ

Tämä tutkielma tarkastelee kuvan ja sanan vuorovaikutusta teknisten tekstien käänämisessä. Tutkielmassa analysoidaan sekä empiiristä että teoreettista aineistoa. Empiirisissä analyyseissä tutkitaan, kuinka joukko maisterivaiheen käännejäopiskelijoita käänsi kuvitettun teknisen tekstin englannista suomeen. Analyysien aineisto koostuu opiskelijoiden käänöksistä ja käänöskommenteista, jotka opiskelijat kirjoittivat käänöstyönsä aikana. Teoreettisissa analyyseissä vertaillaan kognitiivisia käänöstutkimuksia ja kuvitettujen tekstien tulkkintaa käsitteleviä kognitiivisia tutkimuksia, sekä arvioidaan kuvan ja sanan suhteesta muodostettuja teorioita vertaamalla niitä empiirisestä aineistosta muodostettuihin havaintoihin.

Tutkielmalla on kaksi keskeistä tavoitetta. Näistä ensimmäinen on valottaa kuvallisen informaation roolia teknisten tekstien käänämisessä. Tutkimuksessa tarkastellaan, kuinka kuvan ja sanan vuorovaikutusta teknisten tekstien käänämisessä voidaan kuvata teoreettisesti, kuinka opiskelijat itse kuvalevat tästä vuorovaikutusta ja kuinka heidän käänösratkaisunsa mahdollisesti heijastavat tästä vuorovaikutusta. Tutkimuksen toinen keskeinen tavoite on osoittaa, että kuvitettujen tekstien käänöksiä tutkimalla voidaan saada uutta tietoa kuvan ja sanan vuorovaikutuksesta myös yleisemmällä tasolla: käänösten tutkimus tarjoaa mahdollisuuden tarkastella multimodaalista merkityksensrakentumista empiirisesti. Molemmat tutkielman empiiriset aineistot heijastavat käänijien tulkkintoja multimodaalisesta tekstistä. Käänöskommentissa käännejä pohtii syitä ratkaisujensa takana, ja kommenttia analysoimalla voidaan pohtia, kuinka käänijän tulkkinta tekstistä rakentui. Myös käänös heijastaa käänijänsä tulkkintaa tekstistä (esim. Muñoz Martin 2010, 175–176). Näin ollen voidaan ajatella, että multimodaalisen tekstin käänös kuvastaa käänijän tulkkintaa useamman eri moodin yhdistelmästä.

Multimodaalisuustutkimuksessa on useimmiten sovellettu teoreettisia ja menetelmällisiä lähestymistapoja, jotka perustuvat kielitieteeseen: niissä kaikkia moodeja analysoidaan kielitieteellisten käsitteiden avulla. Tämä tutkielma tarjoaa multimodaalisuuden tarkasteluun vaihtoehtoisen teoreettis-analyyttisen viitekehyn, jota tässä työssä kutsutaan *kognitiivisesti orientoituneeksi* lähestymistavaksi. Kielitieteellisesti orientoitunut multimodaalisuustutkimus

tarkastelee kuvan ja sanan vuorovaikutusta peilaamalla sitä teorioihin, jotka selittävät verbaalisen kielen rakentumista. Kognitiivisesti orientoitunut multimodaalisuustutkimus puolestaan tarkastelee kuvan ja sanan vuorovaikutusta peilaamalla sitä siihen, mitä kognitiotutkimus on osoittanut näiden moodien vuorovaikutuksesta kuvitettujen teknisten tekstien tulkinnassa. Lähdestymistavan lähtökohtana on, että kognitiotutkimusten perusteella voimme olettaa näiden tekstien lukijoiden integroivan verbaalista ja visuaalista informaatiota luentansa aikana, ja että tästä syystä lukijoiden tulkinnoja tarkastelemalla voidaan tehdä päätelmiä multimodaalisen merkityksensä rakentumisesta.

Tutkielma koostuu johdanto-osiosta sekä neljästä aiemmin julkaistusta tieteellisestä artikkelistä. Ensimmäisessä artikkelissa muotoillaan teoreettinen viitekehys, joka kuvaa, kuinka käännejä voin olettaa rakentavan tulkintansa kuvitettusta teknisestä tekstistä. Toinen artikeli esittää käänneskomenttien analyysin, jonka tarkoituksena on selvittää, kuinka erilaisin tavoin käännejä opiskelijat käsitteellistivät verbaalisen ja visuaalisen informaation vuorovaikutusta käännesprosessinsa aikana. Artikkelin menetelmänä on fenomenografinen analysi, joka pyrkii havainnollistamaan, kuinka erilaisilla tavilla ihmiset kokevat, ymmärtävät ja käsitteellistävät kohtaamiaan ilmiöitä. Kolmannessa artikkelissa tutkitaan opiskelijoiden käänöksiä valintaverkkoanalyysin (Choice Network Analysis) avulla. Analyysin tavoitteena on selvittää, perustuvatko opiskelijoiden käänökset pelkästään sanalliseen informaatioon vai kuvan ja sanan yhdistelmästä tehtyyn tulkintaan. Neljäs artikeli asettaa tutkimuksen laajempaan, poikkitieteellisen multimodaalisuustutkimuksen viitekehykseen ja pyrkii osoittamaan käänöstieteellisen tutkimuksen annin multimodaalisen merkityksensä rakentumisen tutkimukselle.

Tutkimuksen tuloksista voidaan päätellä, että kuvitetun tekstin käänämisessä ei ole kyse pelkästään verbaalisen informaation välittämisestä kieletä toiselle. Tutkimuksen empiiriset analyysit osoittivat, että lähdetekstin kuvat vaikuttivat lähdetekstistä tehtyyn tulkintaan useilla tavilla, ja että kuvat osaltaan määrittelyt, kuinka verbaaliset elementit kulloinkin käännettävät. Analyysit osoittivat, että kuvitetun tekstin lukija (käännejä) saattaa pyrkiä integroimaan visuaalista ja verbaalista informaatiota silloinkin, kun kuvat ja sanat ovat keskenään ristiriidassa. Tutkimuksen anti käänöstieteen ulkopuoliselle multimodaalisuustutkimukselle kumpuaa työssä esitetystä uudesta teoreettis-analyytisesta viitekehyksestä ja sen avulla tehdyistä havainnoista. Tutkimus osoittaa, että kuvan ja sanan vuorovaikutus on monisyinen, ennalta-arvaamatona prosessi, jota ei aina ole mahdollista mallintaa etukäteen laadittujen kategorioiden avulla, kuten useissa aiemmissa tutkimuksissa on esitetty.

# ABSTRACT

This dissertation analyzes word–image interaction in technical translation by examining both empirical and theoretical material. The empirical analyses of the study examine how a group of Master’s level translation students translated an illustrated technical text from English into Finnish. Its data consists of the students’ translations as well as the translation diaries the students wrote during the translation assignment. The theoretical analyses of the dissertation consist of comparing cognitive studies of illustrated text comprehension with cognitive studies of translation, as well as evaluating theories of word–image interaction by comparing them to the insights gained during the empirical analysis.

The study has two central aims. First, it aims to shed light on the role of images in technical translation. It examines how word–image interaction in technical translation could be modelled theoretically, analyzes how the interaction is described by translation students, and investigates if and how the interaction is reflected in their translation solutions. By doing so, the study aims to further shift the traditional linguistic focus of Translation Studies towards a more multimodally-informed direction.

The second aim of the study is to show that examining the translation of illustrated texts can inform research into word–image interaction also on a more general level – that research into translation offers one possible means to exemplify multimodal meaning construction in an empirical manner. The two types of empirical data examined in the present study are examples of products that reflect translators’ interpretations of a multimodal text. A translation diary is a document in which the translator reflects on reasons behind the translation solutions, and can therefore be used to shed light on how the translator’s interpretation was constructed. A translation, in turn, always reflects the translator’s interpretation of the text (e.g. Muñoz Martin 2010, 175–176). A translation of a multimodal text hence reflects how the translator interpreted the combination of the modes.

The theoretical and methodological approaches that are usually adopted in multimodally-oriented research are linguistically-inspired: they employ linguistic concepts to the analysis of all of the modes involved. The present study has adopted an alternative theoretical-analytical approach, referred to as a *cognitively-grounded*

*approach* to multimodality. The linguistically-grounded approaches to multimodality describe how meaning is constructed from words and images based on theories of how verbal language functions. A cognitively-grounded approach describes how meaning is constructed from words and images based on what research into human cognition has suggested about the way in which we interpret the combination of words and images in illustrated technical texts. The starting point of the approach is that, based on this research, we may assume that the readers of such texts integrate words and images as they read, and that, for this reason, their interpretations reflect the way in which they interpreted the combination of the modes.

The dissertation consists of a summary and four previously published academic articles. The first article outlines a cognitively grounded theoretical framework that aims to describe how a translator's interpretation of an illustrated technical text is likely to unfold. The second article presents the analysis of the translation diaries, which aims to characterize the variation in the ways in which the participants conceptualized the interaction of verbal and visual information while translating the multimodal text. The research approach adopted for this purpose is phenomenography, which sets out to map the different ways in which various aspects of the world are experienced, conceptualized, and understood by different people. The third article examines the students' translations by means of Choice Network Analysis. The article aims to assess if the translations of the multimodal source text were based entirely on verbal information or on a negotiation of meaning from two different modes, the verbal and the visual. The fourth article situates the study in a wider scope of multimodality and aims to demonstrate the value of translational inquiry in the examination of multimodal meaning construction in an interdisciplinary context.

The study concludes that when dealing with illustrated material, translation should not be thought of as entirely verbal activity. The empirical analyses of the study strongly suggest that images played a significant role in source text comprehension and that images defined the way in which elements of the verbal text were translated. The analyses conclude that readers (translators) can make an effort to map visual and verbal information onto each other even when words and images contest or contradict each other. As for research into multimodality in an interdisciplinary context, the contributions of the study stem from working towards establishing novel theoretical and analytical approaches to multimodality. The study argues that word–image interaction is an unpredictable process that cannot always be modelled based on ready-made categories, as is done in several previous studies.

# TABLE OF CONTENTS

PART I .....	16
1      Introduction .....	17
1.1     Background of the Research Problem .....	20
1.2     Objectives.....	22
1.3     Dissertation Structure .....	24
2      Theoretical Background .....	27
2.1     Definitions of Key Terms .....	28
2.2     Theoretical and Analytical Approaches to Multimodality .....	32
2.2.1     The Systemic Functional Approach to Multimodality.....	33
2.2.2     Critique to Systemic Functional Procedures .....	35
2.2.3     Cognitively-Grounded Perspective to Multimodal Meaning Construction.....	39
2.3     Translation Process Research .....	42
3      Data Collection.....	45
3.1     Design of the Source Text.....	45
3.2     Implementation of the Translation Assignment.....	47
3.3     Post-Translation Group Meeting.....	49
3.4     Research Participants .....	51
3.5     Translation Diaries as Research Data.....	53
4      Analysis Methods in the Empirical Articles.....	55
4.1     Phenomenographic Analysis of the Translation Diaries.....	55
4.1.1     Phenomenographic Inquiry.....	56
4.1.2     Evaluation of the Research Approach.....	59
4.1.3     Procedure of Analysis.....	62
4.2     Choice Network Analysis of the Translations .....	64
4.2.1     Choice Networks: Choosing from Options .....	65
4.2.2     Creating a Choice Network .....	67
4.2.3     Interpreting a Choice Network.....	70

5	Analysis of Previous Theories of Word–Image Relationships .....	72
5.1	Modelling Word–Image Interaction in Previous Research.....	72
5.1.1	Word–Image Interaction as Logico–Semantic Relations .....	73
5.1.2	Word–Image Interaction as a Taxonomy of Relationships .....	75
5.2	Procedure of Analysis for the Comparison of Theoretical and Empirical Material.....	77
PART II .....		79
6	Results .....	80
6.1	Cognitively Modelling the Translation of Illustrated Technical Texts.....	80
6.2	Conceptualizing Word–Image Interaction during Translation .....	82
6.3	Examining Word–Image Interaction in the Translations .....	83
6.4	Translation Research as a Means to Model Word–Image Interaction.....	84
7	Discussion.....	87
7.1	Theoretical and Practical Contribution .....	87
7.2	Limitations of the Study.....	89
7.3	Implications for Further Research .....	92
REFERENCES.....		95
PART III .....		112

## LIST OF FIGURES

Figure 1. The mutual relationships between the four articles of the dissertation.

Figure 2. The first image of the source text used in the translation assignment of the study.

Figure 3. An example choice network from the analysis of the data, representing the translations of the prepositional phrase “in the upper part of the equipment”.



# LIST OF ORIGINAL ARTICLES

This dissertation is composed of a summary and the following four original publications, reproduced here by permission. The original publications are referred to as Articles 1 to 4 in this summary.

- 1 Ketola, Anne. 2016. "Towards a multimodally oriented theory of translation: A cognitive framework for the translation of illustrated technical texts." *Translation Studies* 9 (1), 67–81. (Accepted manuscript of article published by Taylor & Francis in *Translation Studies*, available online: <https://www.tandfonline.com/doi/full/10.1080/14781700.2015.1086670>)
- 2 Ketola, Anne. 2015. "Translation diaries of an illustrated technical text. Translation students' conceptions of word-image interaction." *Connexions Special issue: Translation and international professional communication: Building bridges and strengthening skills* 3 (2), 13–40.
- 3 Ketola, Anne. 2016. "An illustrated technical text in translation. Choice Network Analysis as a tool for depicting word-image interaction." *trans-kom: Zeitschrift für Translation und Fachkommunikation* 9 (1), 79–97. (Available online: [http://www.trans-kom.eu/bd09nr01/trans-kom\\_09\\_01\\_06\\_Ketola\\_CNA.20160705.pdf](http://www.trans-kom.eu/bd09nr01/trans-kom_09_01_06_Ketola_CNA.20160705.pdf))
- 4 Ketola, Anne. 2017. "Using translation research to model word–image interaction." *New Voices in Translation Studies* 17: 82–104.

# PART I

# 1 INTRODUCTION

Many of the texts we translate today include both words and images, and hence communicate by combining verbal and visual information. Yet, translation is often understood as a predominantly verbal activity; as Ira Torresi (2007, 52) writes, the role of images in translation largely remains *terra incognita* in Translation Studies. Maeve Olohan (2016, 55) has pointed out that images are often not changed in technical translation. This might be the reason why the role of images in technical translation is often ignored (Kokkola and Ketola 2015). My dissertation starts from the premise that even if images do not physically change in translation, they may still affect translation in various ways.

This dissertation examines word–image interaction in technical translation by analyzing both empirical and theoretical material. The dissertation consists of the present summary and four previously published scientific articles, two of which involve theoretical analysis (Articles 1 and 4) and two involve empirical analysis (Articles 2 and 3). The empirical parts of the study analyze how a group of eight Master’s level translation students translated an illustrated technical text from English into Finnish. The empirical data consists of the students’ translations as well as the translation diaries the students wrote during the translation assignment. The first part of the theoretical analysis consists of comparing cognitive studies of illustrated text comprehension with cognitive studies of translation in order to outline a cognitively grounded theoretical framework that aims to describe how a translator’s interpretation of an illustrated technical text is likely to unfold. The second part of the theoretical analysis consists of re-evaluating theories of word–image interaction by comparing them to the insights gained during the empirical analysis.

The study could be placed in the intersection of Translation Studies and the rapidly expanding, interdisciplinary field of research into multimodality, and it attempts to break new ground on two fronts. First, it aims to shed light on the role of images in technical translation. It examines how word–image interaction in technical translation could be modelled theoretically, analyzes how the interaction is described by translation students, and investigates if and how the interaction is reflected in their translation solutions. Second, the study aims to elucidate how

research into the translation of illustrated texts could inform research into word–image interaction on a more general level. As affirmed by, for instance, Muñoz Martin (2010, 175–176), translators do not transfer texts from one language to another; they transfer their own *interpretations* of these texts. If we examine the translation of a multimodal text, we are examining how a particular translator interpreted the combination of the modes involved. Comparing a set of translations of an illustrated text can exemplify the multitude of ways in which the combination of words and images can be interpreted. Studying the process of translation may therefore offer valuable insights into how we construct meaning from different types of texts.

The empirical data of the study was produced by a group of Master's level translation students who translated an illustrated text. This text presented the operating principles of two different types of ore beneficiation devices used in the mining industry (see Appendix 1). In August 2013, I visited the Eastern Finland office of the Geological Survey of Finland (GSF) in person and examined the ore beneficiation devices in action. The source text used in this study was created during this two-day visit with the assistance of GSF staff. Their eagerness to assist me in the research project, I was told, was due to the fact that they often struggle to find translators with experience in their specialized subject domain, and therefore found it important to promote research addressing mining-related translation challenges.

During this research project, I have received three comments about my topic that have further convinced me of the importance of my research. The two first of these comments were given by technical translators, and the third by a technical communicator. The first of the technical translators told me that images are not involved in the work of technical translators like herself who work with translation memory programs. She admitted that nearly all of her translation commissions include illustrations, but as only the verbal part of the source text is fed to the translation memory, the images are not a part of her translation practice. She then continued that she briefly looks at the image when faced with a name of a part of a device – to see what the part is – but then goes back to work on the verbal dimension of the source text. While, for the translator, this meant that the image is not involved in translation, I take this to reflect the opposite. I propose that when dealing with technical texts, *we tend to rely on images to define the contents of words*.

The second comment was given by a technical translator who told me that while the images are right next to the verbal text he is translating, he almost never looks at them. He then added that he only tends to look at the image when there is something wrong with the image. He explained that his clients sometimes print the manual of

a new model of a device with the images of the previous model (which, according to the translator, is not uncommon). I propose that this is not a question of not looking at the images while we translate, but of not paying conscious attention to looking at the images – how else would we know when there is something wrong with the image and when not? My point is that *we might not be aware of how just how much we look at the images we encounter*.

The third one of these comments was given by a technical communicator who told me that in the technical communication company she works for, they include images in almost every document produced but not much attention is paid to their selection. She asserted that the company has a large database of images of varying degrees of specificity and technicality, and that images are often selected for documents based on their visual appearance. They might select an image that might nicely fit the color scheme of the document or fit the space that is left blank after adding verbal text on a page. While producing visually appealing documents is undoubtedly important, what this practice does not take into account is that the images might well affect the way in which the verbal text is interpreted. Eye-tracking research in the field of Educational Psychology has affirmed that when reading technical texts, readers make frequent switches between the words and the images, reflecting a process of constructing a joint interpretation of verbal and visual information (e.g. Mayer and Gallini 1990; Hegarty and Just 1993). What this means for those who read the documents produced by this particular technical communication company is that the reader might actually make an attempt to interpret the verbal text in relation to the almost randomly selected image. While this document production practice is undoubtedly not common to all technical communication companies, the example goes to show that *not even the producers of illustrated texts might always be aware of effect words and images can have on each other*.

In the remaining part of this introductory chapter, I present the design of my research project. I start by describing the context of my study and discussing why research on the subject is needed. I then introduce the objectives of the dissertation: I define my research questions and describe the perspective from which each of them approaches the subject. Finally, I introduce the structure of the dissertation and describe how the four articles of the study interrelate.

## 1.1 Background of the Research Problem

Multimodality refers to the coexistence of more than one mode – such as written language, spoken language, and images – within a particular context (Gibbons 2012, 8). Multimodally-oriented research stems from the idea that all modes of a multimodal text contribute to the overall meaning-making potential of the text (Jewitt 2009a, 14). While the analysis of verbal language is often an important part of the investigation of multimodal texts, multimodally-oriented research posits that verbal language is always embedded within – and interpreted in relation to – a wider multimodal context (*ibid.*).

When reading an illustrated text, readers process both verbal and visual information, words and images, and form their interpretation of the multimodal text based on information provided by both of these modes. This idea has been established in various disciplines, examining various types of readers – from elementary school students to educated grown-ups – and various types of illustrated texts – from scientific texts to children’s picturebooks. The methods employed in these studies have ranged from eye-tracking (e.g. Holsanova, Holmberg and Holmqvist 2009) to reading comprehension tests (e.g. Mayer 2002; Schnotz 2005) and reader interviews (e.g. Connors 2013; Youngs and Serafini 2013). This study sets out to test the assumption that since translators of illustrated texts start their work as readers, the same argument applies for them as well – that translators form their interpretation of the illustrated source text based on a combination of verbal and visual information. This study aims to shed light on what the implications of this might be for the actual translation solutions.

The consideration of multimodal issues has been steadily gaining importance in Translation Studies in recent years; after all, a large number of the texts being translated today are multimodal (Hirvonen and Tüttula 2010, 1). Multimodality in translation gained its own entry in the *The Routledge Handbook of Translation Studies* in 2013, written by Klaus Kaindl, as well as *A Companion to Translation Studies* in 2014, written by Luis Pérez-González. As to examining the interaction of verbal text and still images in translation, it might be safe to suggest that the most prominent research has been conducted in regard to children’s picturebooks, most notably by Riitta Oittinen (e.g. 1990, 2003, 2004, 2008). In recent years, others have followed her lead (e.g. Van Meerbergen 2010; Garavini 2014; Martínez-Mateo 2014). Other studies have addressed the role of images in the translation of advertisements (e.g. Torresi 2007, Damaskinidis 2015), magazines (Chueasuai 2013) and comics (e.g. Kaindl 2004). The research designs in these studies have differed from the design

implemented in my research both in regard to the subject matter as well as the nature of the data: they have compared an illustrated source text to its (single) translation. To the best of my knowledge, this dissertation is the first study in the discipline to collect a set of translations produced by multiple research participants and analyze it from the perspective of word–image interaction.

In this dissertation, technical translation is understood as the translation of texts involving or concerned with applied and industrial sciences (see Section 2.1). Research into technical translation from a multimodally-oriented position is of great importance for two reasons: Firstly, technical translation is a significant area of translation, constituting a major share of the global translation market (Byrne 2012, 6). Secondly, different types of images are a key feature of technical texts (e.g. Risku and Pircher 2004; Tercedor et al. 2009, 143; Byrne 2012, 26, 54; Olohan 2016, 54). Therefore, most technical translation commissions will involve visual information in addition to the verbal.

Despite the prevalence of images in technical texts, fairly little research had been conducted on the role of images in technical translation. As Tercedor et al. write,

[i]mages are often part of the textual structure of Scientific and Technical texts. The understanding and interpretation of the text often depends on the way the interface between the text and the image is established. Despite the key role images play in specialised translation, there is a lot of work to be done to increase awareness about the role images play in communicating science and technology (2009, 143).

The translation of illustrated technical texts has previously been studied in detail mainly from the perspective of appropriate image selection. Tercedor-Sánchez and Abadía-Molina (2005) examine the need to replace the images of the source text during translation in order to ensure their proper transmission to a target audience, and provide criteria for choosing target audience-appropriate images in technical and scientific texts (for discussion on similar issues within research into localization, see e.g. Horton 2005; Hiippala 2012). Regardless of the importance of these considerations, the reality of technical translation practice is that images are rarely changed in translation (Olohan 2016, 55). Research is therefore needed on how the original source text image is involved in the translation.

Research in the field of psychology has examined the reading process of illustrated technical texts for over two decades. As early as 1993, Mary Hegarty and Marcel A. Just conducted a study in which they monitored the eye-fixations of test subjects reading an illustrated technical text. They found that the readers frequently interrupted their reading of the verbal text to inspect the image (an average of 6 times per a page of 140 words). Instead of looking at the image at random, the readers

fixated on the parts of the image that depicted objects and events they had just read about (Hegarty and Just 1993, 730–731). Subsequent research in psychology has set out to create cognitive models that explain what exactly occurs in the mind of the reader when they switch their visual attention in this fashion. No consensus exists as to how verbal and visual information is integrated in the mind of the reader: some posit that the reader constructs a single, joined representation of the two modes (e.g. Mayer 2005), others affirm that various competing representations are created of each mode (e.g. Schnottz 2005). Yet, both views agree on the assertion that each mode profoundly affects how the other is interpreted. From the perspective of translation, this means that visual information is able to shape translators' interpretation of verbal material.

The overall research problem of my dissertation can be described as *how do words and images interact during the translation of an illustrated technical text?* In addition to shedding light on multimodal meaning construction in translation, my research sets out to examine multimodal meaning construction on a more general level. As Paul Kussmaul describes, research into translation can be used to examine what goes on in the human mind on a more general level: Kussmaul posits that a translation can be examined as a free production experiment – albeit in a different language – which reflects a person's interpretation of a particular text (2000, 69). I have therefore set out to examine what research into the translation of multimodal material can tell us about the way in which we process this material also in a more general setting.

## 1.2 Objectives

The study set out to investigate how translation students build their interpretation of a multimodal text consisting of visual and verbal information and to develop a deeper understanding of the negotiation of meaning that takes places between verbal and visual information. The empirical data of the study consisted of eight translations of an illustrated technical text, from English into Finnish, produced by Master's level translation students, as well as the translation diaries the students wrote during the translation assignment. The study set out to answer four research questions, each of which is covered in a separate journal article. The question numbers correspond to the article numbers introduced in the following section, as well as in the beginning of this dissertation (see List of Original Papers). The research questions were formulated as follows:

1. *How can the translation of an illustrated technical text be cognitively modeled based on cognitive studies of illustrated text comprehension?*
2. *How do translation students conceptualize the interaction of verbal and visual information when translating an illustrated technical text?*
3. *Do the students' translation solutions reflect word–image interaction? If so, how?*
4. *How can translation research be used to advance our understanding of multimodal meaning construction outside the immediate context of Translation Studies?*

As mentioned above, my overall aim has been to examine how words and images interact during the translation of an illustrated technical text. The first three research questions approach this aim from three different perspectives: examining the process theoretically (*What may be expected to happen during translation?*), examining the participants' own accounts of translation (*What do the translation students report happening?*), and examining the translations (*What appears to have actually happened?*). The fourth research question explores how the examination of word–image interaction in translation can promote our understanding of multimodal meaning construction on a more general level.

In addition to these more particular goals, my research aims to further promote the multimodally-driven perspective into translation, witnessed in the discipline in recent years. Consolidating novel perspectives in research inevitably also brings about the need for both theoretical as well as methodological development. The theoretical positions that are usually endorsed in research into multimodality tend to rely on linguistic paradigms: for instance, the social semiotic and the systemic functional approaches into multimodality are both based on M.A.K. Halliday's social systemic functional grammar (Jewitt 2009b, 28–39). In these approaches, images are treated as structurally equivalent to verbal language. This dissertation is an attempt at adopting a view of multimodality outside these positions.

As Tuominen et al. (2016, 13) emphasize, translation research requires further development of different methodological approaches to address multimodally-oriented research questions. The analysis methods employed in this dissertation are phenomenographic analysis for the examination of the translation diaries and Choice Network Analysis for the examination of the translations. My research is an effort at methodological experimentation, as neither of these methods has been previously used to address multimodally-driven research questions.

My research also aims to emphasize the importance of including multimodal considerations in translator and technical communicator training. As Hanna Risku and Richard Pircher conclude (2008, 165), paying attention to visual information is an increasingly important part of the work of translators and technical communicators and it is a skill these professionals need to acquire. The research participants examined in my study are Master's level translation students, so the observations made in the study are particularly enlightening for examining multimodal source text analysis skills in the context of developing translation competence.

Finally, my study aims to make an empirical contribution to the interdisciplinary field of multimodally-oriented research. Word–image interaction has been the subject of an abundant number of studies in the last decades in a variety of disciplines. Yet, John A. Bateman (2014, 39–40) remarks that the observations made about word–image interaction are generally empirically unsubstantiated and seldom move beyond providing theoretical assumptions. My study proposes that research into translation may offer one possible way to empirically exemplify how the combination of words and images is interpreted.

### 1.3 Dissertation Structure

This dissertation consists of three parts. Part I introduces the theoretical background of the study as well as the data and the methods used in the empirical parts. Part II of the dissertation summarizes the main findings of the study. Finally, Part III includes the four original articles that form the main contents of this dissertation.

Article 1 forms the theoretical basis of the dissertation. It outlines a cognitively grounded theoretical framework for examining the translation of illustrated technical texts by analyzing what has been proposed about translation as cognitive activity, and comparing this with cognitive studies of illustrated text comprehension. The translation process involves various cognitive activities, including source text comprehension, transfer between the two languages, and the production of the acquired information in the target language (Englund Dimitrova 2010, 406). Article 1 does not examine the translation process as a whole, but focusses on the processes involved in source text comprehension and their result: in Jensen's (2010, 216–217) terms, *the mental representation* constructed of the source language message. This part of my dissertation proposes that the mental representation a translator constructs

from an illustrated source text is often negotiated from the combination of verbal and visual information.

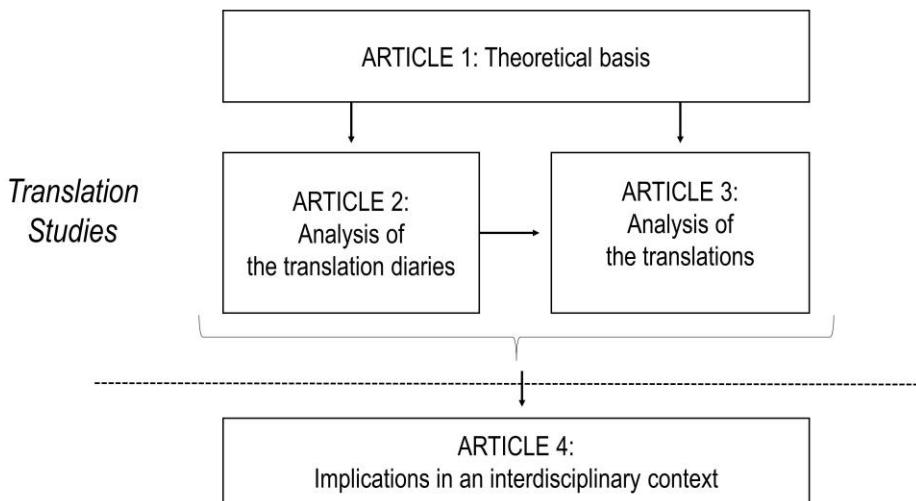
Article 2 presents a phenomenographic analysis of the translation diaries. Phenomenography aims to identify the different ways in which we conceptualize phenomena in the world around us. The aim of Article 2 is to characterize the different ways in which the translation students conceptualized the interaction of verbal and visual information while translating the multimodal text. In the article, I propose that the data reflected two main ways of conceptualizing the interaction of verbal and visual information: either conceptualizing the combination of modes as an entity to be perceived as a whole, or conceptualizing the modes as competing sources of information. All in all, the article suggests that visual information was integrally involved in translation: the students did not assign it to a merely decorative role but, instead, conceptualized it as an important part of the source text.

Article 3 examines the translations by means of Choice Network Analysis. The method comprises comparing the translations of the same source text by multiple translators in order to empirically derive the options – the set of possible solutions – that were available when translating each verbal item. The article aims to assess if the options offered by the multimodal source text were based entirely on verbal information or on a negotiation of meaning from two different modes. In this part of my dissertation, I argue that the image is capable of reattributing the meaning of a verbal element. In the most extreme cases, conflicting visual information may cause verbal information to be disregarded altogether.

Article 4 discusses how the empirical analyses of my study, introduced in Articles 2 and 3, can be used to offer an empirical illustration of the multitude of ways in which the combination of words and images can be interpreted. The article hence explores how the observations made in the study can inform research into word–image interaction outside the immediate context of translation. The article critically reviews two studies which present a classification of the possible ways in which words and images can co-construct meaning, and compares these to the empirical insights gained in Articles 2 and 3. With the help of the empirical examples, I argue that multimodal meaning construction is in fact more complex than the reviewed classifications suggest and that conceptualizing word–image interaction by modelling it on simplified classifications may limit the way we perceive the complexity of this interaction.

Figure 1 presents how the four articles interrelate. Three first three articles are directly situated in the scope of Translation Studies; the fourth reaches outwards from the immediate context of translation and situates the entire research project in

a wider scope of multimodality. Article 1 sets out the theoretical rationale for the rest of the articles. The analysis of the translation diaries, presented in Article 2, is used to inform the analysis of the translations in Article 3. Article 4 reflects on the implications of the study for the examination of multimodal meaning construction on a more general level.



**Figure 1.** The mutual relationships between the four articles of the dissertation.

In the remainder of Part I of the dissertation, Chapters 2 to 5, I introduce the background of my research. In Chapter 2, I discuss the definitions of the most relevant concepts of the study, present my theoretical approach to multimodality, and discuss how my research relates to process-oriented translation research, with which my research shares points of contact. In Chapter 3, I present the empirical data of the study and, in Chapter 4, I introduce and reflect on the methods used to analyze it. Chapter 5 discusses the theoretical material on word–image interaction I examine in Article 4 and introduces the procedure of analysis I followed when comparing these theories to my empirical work.

Part II of the dissertation consists of two chapters: Chapter 6 discusses the main research findings made in my study and Chapter 7 evaluates the contributions of the study, discusses its limitations and provides recommendations for future research. The original articles of the dissertation are reprinted in Part III of the dissertation.

## 2 THEORETICAL BACKGROUND

This chapter provides the definitions of the key concepts used in my research and situates the study in a theoretical context. Section 2.1 is dedicated to introducing the key concepts employed in the study. I begin the discussion by reflecting on the definitions of *technical translation* and *technical text*. I then analyze the components of a multimodal text, starting with definitions of more particular terms – *images* and *words*, *verbal* and *visual (source) texts* – and then move on to reflect on the more general terms, namely *mode* and *multimodality*, as well as the *interaction* of modes and *word–image relationship*.

Section 2.2 discusses ways in which multimodality has been approached theoretically and analytically in previous research and how it is approached in the present study. I start by reviewing one of the most commonly adopted approaches to multimodally-oriented research, in Translation Studies as well as other areas of research, namely the systemic functional approach to multimodality. This approach belongs to an area of research that could, perhaps, be described as *multimodality in the humanities*. I then move on to describing my own approach which, in turn, could be broadly situated in an area of research described as *multimodality in psychology*. In Section 2.2.1, I present an overview of the underlying principles that inform the systemic functional approach to multimodality. In Section 2.2.2, I review two previous applications of the approach in the field of Translation Studies and discuss how the underlying principles of the approach affect – and possibly complicate – analysis on a practical level. In Section 2.2.3, I discuss how the conceptual basis of my study differs from this approach and introduce the research that informs the theoretical rationale behind my own perspective to multimodality. Finally, Section 2.3 discusses translation process research and the way in which the present dissertation differs from research designs typically adopted in process-oriented studies.

## 2.1 Definitions of Key Terms

In this dissertation, *technical translation* is understood as the translation of texts involving or concerned with applied and industrial sciences. This definition is derived from the way in which the Oxford Dictionary of English (2010) defines the word *technical*. The definition of technical translation adopted in this dissertation hence largely differs from others that view technical translation as synonymous with the translation of all texts written using Languages for Special Purposes (LSP). For instance, in their introduction to the book *Scientific and Technical Translation* (1993, 1), Sue Ellen Wright and Leland Wright define technical translation as encompassing translation in subject domains ranging from engineering to economics, psychology and law. Such an open approach to technical translation is justified if technical translation is to be viewed, above all, as an activity requiring a thorough understanding of a particular subject domain and knowledge of its terminology. However, this dissertation needs to employ a stricter definition of technical translation because I build many of my arguments on research on illustrated text comprehension that specifically addresses texts concerned with understanding applied and industrial sciences.

In academic discussion, technical translation is often paired together with scientific translation, as also reflected in book titles such as Jody Byrne's *Scientific and Technical Translation Explained* (2012), Maeve Olohan's *Scientific and Technical Translation* (2016) or Wright and Wright's *Scientific and Technical Translation* (1993) discussed above. As Byrne (2012, 2) affirms, the lines separating scientific and technical texts are increasingly blurred. Olohan (2016, 6–7) proposes that dictionary definitions of science and technology seem to suggest that technology emerges from science as its practical application. However, she notes that the primacy of science and the subordination of technology can also be disputed in the technology-driven era we live today. Byrne (2012, 2) observes that it is common for texts to include elements of both scientific and technical translation. This is also the case with the source text used in the present study (see Section 3.1): it is designed to explain how a device works, but it is also designed to discuss and explain the scientific principles behind the operating principle.

Byrne (2012, 26) describes *technical texts* as documents that seek to help an audience understand a subject, understand how to carry out procedures efficiently and how to avoid accidents when using something. When discussing *illustrated technical texts*, this dissertation refers to informative, instructive texts that explain how something works by both verbal and visual means, in other words, by employing

both words and images. Describing the object of my study as *illustrated texts* comes with the risk of foregrounding the verbal in relation to the visual. The term can be interpreted to reflect that visual information is merely a property that characterizes the verbal text. Despite this possible limitation, I have opted to use this term for the lack of other, reasonably explicit and concise options. The term *multimodal* text, as discussed below, covers a variety of multimodal combinations of information, many of which my discussion does not cover. Referring to the object of my study as “texts that include both verbal and visual information” throughout the dissertation, in turn, seems unnecessarily complicated.

In order to emphasize the integral role of visual information in my research, I have chosen not to refer to visual information as *illustrations* but as *images*<sup>1</sup>. The two words, together with *picture*, may be considered as near-synonyms. In everyday use at least, one could claim that the three are often used interchangeably. The word *picture* is derived from the Latin *pictura*, meaning “painting”. From an etymological point of view, the word hence better describes a visual product that has been made with artistic purposes or one that carries some type of aesthetic relevance. The word *illustration*, in turn, derives from late Middle English via the Latin verb *illustrare* (“illuminate”), which connotes the idea of providing an example of something; in the case of illustrations in the form of visual information, this would mean providing an example of verbal information. Describing a visual product as an *illustration* therefore implies that the artifact has been made specifically to accompany verbal information of some sort. While, strictly speaking, this is true for the visual products produced for the purposes of this research project, I feel that employing the term downplays the importance of visual information. These visual products can also serve as sources of information in their own right as opposed to merely providing visual examples of verbal information. Etymologically, *image* is the most versatile of the three. It derives from the Latin *imaginem* (nominative of *imago*) which can refer to a variety of concepts, including “depiction” (in an artistic sense), “copy” or “imitation”, “idea” or “appearance”, and “conception” or “thought” (Merriam-Webster). Out of the three near-synonyms, *image* has the most potential to represent a concept that has the ability to stand on its own as a representation of information and that is not necessarily an artistic expression of any sort.

In this study, I examine *technical images*. I emphasize that the observations I make about the role of images in technical translation are not likely to apply, as such, to

---

<sup>1</sup> In the published articles featured in the dissertation, the actual images of the data are referred to as *figures*. This has been entirely due to the conventions of academic writing: most academic journals refer to all types of images as *figures*.

the translation of other text types combining words and images, such as advertisements or comic books. The reason for this is that different types of images are likely to affect our illustrated text comprehension in different ways (Mayer 1993). As I discuss in Section 7.3, a possible, intriguing line of future research would be to perform similar analyses on translations of other illustrated text types, and compare these analyses.

In multimodally-oriented (translation) research, the term *visual* is used to refer to two different things. First, it can refer to information that is perceived with *sight*: for instance, research into audio description needs to make the distinction between the information perceived visually (eyes) and aurally (ears) (e.g. Hirvonen 2012; Jimenez Hurtado and Soler Gallego 2013). Second, the term can refer to *pictorial* information (e.g. Pereira 2008; Oittinen 2008). In this study, I use the term *visual* to refer to information that is presented in a pictorial form, and *visually perceived* to refer to information perceived with sight. This takes us to the next definition: in this dissertation, *verbal text* refers to a visually perceived instance of language use. In other words, the definition of verbal text in this study does not include language use that is perceived auditively, for reasons discussed further below. In a similar fashion, the term *word* in this study refers to words that have been visually perceived. I have chosen to refer to verbal language use as *verbal text* as opposed to *text* only, in order to clearly distinguish it from my other uses of text.

In Translation Studies, the source information on which a translation is based is typically referred to as a *source text*. As the aim of this study is to examine the role of images as an integral part of this source information, I have chosen to divide the source text into two components: I refer to the verbal part of the source information (the words) as the *verbal source text* and the visual part of the source information (the images) as the *visual source text*. I admit that the notion of a *visual text*, in itself, can be easily contested. One of the central aims of multimodally-oriented research is to shift our focus from the dominance of verbal language as a research object and admit the importance of other modes as resources for meaning making. Referring to images as visual texts can be said to describe them from the perspective of verbal language, and encourage one to examine images as if they had the properties of verbal language – as is done, for instance, by Gunther Kress and Theo van Leeuwen in their *Reading Images: The Grammar of Visual Design* ([1996] 2006) and the abundance of studies building on their ideas. Yet, I feel that describing the images of my study as the visual source text also emphasizes their importance in a way: it highlights that they, too, can be the source on which a translation is based.

After examining words and images separately, I will now move on to consider what they can be said to have in common as meaning-making resources by reflecting on the concept of *mode*. This is, perhaps, the most central concept of multimodally-oriented research. Following John Bateman (2014, 18), mode is understood in this study as “some organized bundle of properties held in common by some material that, when used following agreed conventions, provides some community of users with a distinctive way of making meanings with that material.” For images, this bundle of properties would include shapes, (relative) sizes, colors, tonal contrasts, and depth (e.g. Stöckl 2004, 14). In order to be comprehended by a particular community, these properties would have to be arranged in a way that sustains features of structural likeness to some referent which the image aims to depict.

For words, the bundle would include purely linguistic properties such as grammar and lexis as well as different graphic properties such as the use of typefaces and layout (Stöckl 2004, 11). In order to be comprehended, these properties would have to be arranged following the conventions of a particular language and adhering to certain conventions of graphic representation. My definition of verbal text, above, distinguishes between visually and auditively perceived verbal information because they are regarded as separate modes in this study – and, one might propose, most multimodally-oriented studies. The bundle of properties auditively conveyed verbal text draws on are slightly different: in addition to linguistic properties, the bundle would include intensity (loudness), pitch, intonation and so on.

As mentioned in the Introduction section, the term *multimodality* is employed in this study to refer to the coexistence of modes – written language and images in particular – within a particular context, such us as an illustrated technical text (cf. Gibbons 2012, 8). A *multimodal text*, then, here refers to a message created by using the properties of different modes. In this study, I have strived to avoid making the distinction between *verbal* and *non-verbal* modes. As Şebnem Susam-Sarajeva writes in regard to the division of *Western* and *non-Western* research models in Translation Studies, “[a]ny adjective describing its subject as a negation, as a ‘non-x’, is derived from the vantage point of the ‘x’” (2002, 193). Describing a mode as non-verbal, therefore, does not advance a line of thought that aims to treat all modes as equally important in their meaning making potential<sup>2</sup>.

Perhaps the most fundamental concept of this study is *word–image interaction*. A dictionary definition of *interaction*, as proposed by the Macmillan Dictionary (2017), refers to “the process by which different things affect each other or change each

---

<sup>2</sup> This notion was first brought to my attention by my colleague Sari Kokkola. I am thankful for this, as well as the many other inspiring conversations we have had about multimodality.

other”. Building on this general idea, I define word–image interaction as the process by which words and images affect and/or change the way in which the other is perceived in the cognitive system of the reader (translator) of an illustrated text. Following Jana Holsanova (2012, 252), I emphasize the perspective of the reader of a multimodal text in defining this concept: words and images only affect each other when perceived by someone.

## 2.2 Theoretical and Analytical Approaches to Multimodality

As Carey Jewitt (2014, 134) concludes, the origins and current forms of multimodally-oriented research are complex; it may be conducted from a variety of theoretical positions, each of which offers “a particular starting point and pathway” into the manifold subject. In the humanities, the three most commonly adopted theoretical perspectives to multimodality are *multimodal interaction analysis* (focusing on how individuals react in certain situations), as well as the *social semiotic* and the *systemic functional* approaches, both based on the work of the linguist M.A.K. Halliday (for comprehensive reviews on these approaches, see e.g. Jewitt 2009b). As the name suggests, the social semiotic approach focuses on the social context of meaning making: it examines people’s situated use of different modal resources (*idem.*, 29). The systemic functional approach views modes as structured systems that have evolved in accordance to the functions they have in societies (*idem.*, 31–32).

In translation research, there have been various multimodally-oriented studies which set out to compare verbal and visual information in translation without explicitly building on any of these established theoretical approaches to multimodality (e.g. Garavini 2014; Kaindl 2004; Ketola 2016b; Ketola and Martínez Mateo 2018; Liu 2011; Oittinen 2004, 2008). Multimodally-oriented translation research that does specify a theoretical stance to guide the methodology, most often, adopts the systemic functional perspective to multimodality. Examples of this include analyses of subtitles (e.g. Mubenga 2009, 2010; Espindola 2012) and translations of advertisements (e.g. Millán-Varela 2004; Feng and Espindola 2013). The approach has also been employed in two studies examining translations of illustrated texts (Van Meerbergen 2010; Chueasuai 2013).

The prominence of the systemic functional approach to multimodality in translation research is also emphasized in Luis Pérez-Gonzalez’s article “Multimodality in Translation and Interpreting Studies”, written as a general introduction to the topic for *A Companion to Translation Studies* (2014). In the article,

the author discusses the arrival and the promise of “multimodal theory” to translation research. Later in the article, it is explained that this term explicitly refers to the systemic functional and the social semiotic approaches only (Pérez-Gonzalez 2014, 127). Referring to these specific research approaches with a term as general as “multimodal theory” could even be interpreted as suggesting that *multimodality* is synonymous with the systemic functional and the social semiotic approaches.

The systemic functional approach to multimodality has greatly contributed towards our awareness about multimodality in communication and generated a vast and robust area of research interest. However, in this dissertation, I have opted to examine multimodality from a perspective that conclusively differs from this approach. In this section, I present the systemic functional approach in order to justify why introducing an alternative approach is beneficial to the development of multimodally-oriented (translation) research. In Section 2.1.1, I present an overview of the underlying theoretical underpinnings that inform the systemic functional approach to multimodality. In Section 2.2.2, I move on to introduce how the approach has been previously applied when analyzing the translations of illustrated texts and discuss the possible weaknesses of this approach based on these previous studies. Finally, in Section 2.2.3, I introduce my own approach, present the theoretical rationale that underlies it, and discuss the main methodological implications of the approach.

## 2.2.1 The Systemic Functional Approach to Multimodality

The systemic functional approach to multimodality is one of the most influential approaches in all multimodally-oriented research in the humanities. The approach is linguistically-inspired: it is based on M.A.K. Halliday’s (e.g. 1985) theory of systemic functional linguistics and it was developed for a multimodal context most notably by Gunther Kress and Theo van Leeuwen ([1996] 2006, 2001). Kress and van Leeuwen’s work offers a view of multimodality in which “common semiotic principles operate in and across different modes” (2001, 2) and it therefore treats all modes as identical in their meaning-making strategies.

According to the systemic functional approach to multimodality, all modes realize the three metafunctions Halliday (1978) proposed for language: the *ideational*, the *interpersonal* and the *textual* (which is referred as the *compositional* metafunction when discussing images) (Kress and Van Leeuwen [1996] 2006). In the metafunction analysis of verbal language, the ideational metafunction refers to the way in which

external reality is represented in the text. In other words, it refers to the content of the text: the people, places, actions, circumstances and so on, being discussed (e.g. Halliday and Matthiessen [2004] 2013, 29). The interpersonal metafunction refers to the relationship that exist between the author of the text and the reader of the text (*idem.*, 29–30). The author may address the reader in various ways, for instance, by instructing, negotiating, persuading or giving orders. The textual metafunction refers to the property of the text that links individual words into coherent clauses. It involves aspects such as cohesion, information structure and thematic structure (*idem.*, 30).

The metafunction analysis of images is founded on the social semiotic account of how meaning is realized visually, introduced by Kress and van Leeuwen in their seminal *Reading Images. The Grammar of Visual Design* ([1996] 2006). Image analysis on the ideational level includes identifying the *participants* of the image (people, objects, places) as well as events and relationships between the participants (Kress and van Leeuwen [1996] 2006, 42). Image analysis on the interpersonal level refers to identifying the presumed relationships between the creator of the image, its viewer and the participants in the image. The metafunction is said to be realized on three levels: eye contact, size of frame (close shot, medium shot, long shot) and perspective (*idem.*, 42–43, 116–129). Image analysis on the compositional level refers to identifying the composition of the image, the color and the relative size of the participants and their positioning (up, down, left, right, center) in regard to the rest of the participants (*idem.*, 43, 177–179).

The analytical focus of the systemic functional approach is in describing the metafunctionally-based systems for all of the modes under study. The metafunction analysis of an illustrated text, therefore, starts with two separate analyses: identifying the ideational, interpersonal and textual or compositional dimensions of the verbal text on the one hand and the images on the other hand. When analyzing the translation of illustrated texts, the method would typically entail performing a metafunction analysis on the verbal source and target texts, on the visual source text (images of the source text), and – if the images have been modified during the translation process – on the visual target text (images of the target text). In order to make observations about multimodal meaning construction, the analyses of the verbal and visual dimensions should then be combined. In the following subchapter, I move on to discuss how this analysis procedure has been carried out in the two studies examining word–image interaction in the translation of illustrated texts.

## 2.2.2 Critique to Systemic Functional Procedures

The systemic functional approach, or the metafunction analysis, has been employed to analyze the translations of illustrated texts by two authors. Pasakara Chueasuai's (2013) study examines the Thai translations of the *Cosmopolitan* magazine, originally published in the United States. Chueasuai's research examines how representations of sex and sexuality have changed when the magazine is brought from the sexually liberated source culture to the more conservative target culture (2013, 107–108). In addition to translating the articles of the magazine from English into Thai, the Thai publisher has replaced some of the magazine's illustrations. Sara Van Meerbergen's (2010) study examines the Swedish translations of Dick Bruna's *Nijntje* picturebooks, originally published in Dutch. The study analyses how the books' verbal and visual resources are used in the source and target texts on different metafunction levels (for a more thorough comparison of these studies, see Ketola 2016a; not a part of this dissertation).

The systemic functional tools of image analysis are often complimented for offering *systematic* procedures for analyzing multimodal material (e.g. Martinec and Salway 2005, 341). However, it is my impression that these analyses are not always carried out in a systematic manner, which would mean analyzing both modes on all three metafunction levels. The textual-compositional metafunction level, for instance, did not seem to offer a useful starting point for analysis for either of the authors. Chueasuai's study does not address the metafunction level at all, and Van Meerbergen's textual-compositional analysis, as I have argued in more detail elsewhere (Ketola 2016a, 100), focuses on which objects or participants are named in the verbal source text and the verbal target text, and hence can be suggested to operate on the ideational rather than the textual metafunction level. Moreover, both studies include various examples of analysis being conducted on one mode but not the other. For instance, in Chueasuai's study, images have only been analyzed on the ideational level.

Perhaps more importantly, it seems that the metafunction analyses of verbal language and images, performed separately, cannot always be combined in a productive manner. If the analyses cannot be combined, what we are discussing is a procedure for producing separate, *monomodal* analyses instead of a multimodal analysis of the interaction of the modes. In Chueasuai's study, the analyzed segments of verbal and visual information did not even correspond to same magazine issues, so combining the analyses would have been impossible. The analysis offers separate examples of changes in the verbal and in the visual dimension of the material. The

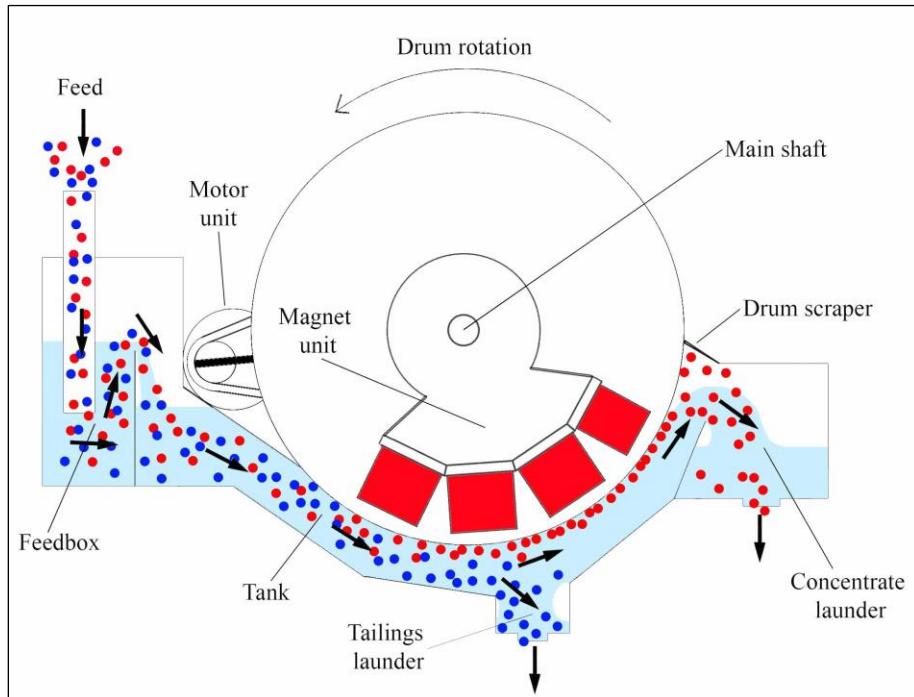
author comments on this by concluding that the analyses are presented separately “for ease of discussion and clarity in the presentation of the results” (Chueasuai 2013, 118), but no suggestions are offered as to how the two stages of analysis *could* be combined.

Both of the reviewed studies offer interesting insights into how illustrated material may be handled in translation, but I argue that their procedures of analysis, apart from the few exceptions, do not address how meaning is constructed multimodally. As I propose elsewhere (Ketola 2016a, 112–113), metafunction analyses often produce detailed lists about the properties of individual modes, but generally do not seem to advance beyond itemizing the metafunctionally-based structures for both of the modes separately. Charles Forceville (2007, 1236), too, has discussed the problem of “infinite detail” in multimodal research: many of the analyses produce painstakingly detailed descriptions of the components of multimodal texts but rarely formulate generalizations about multimodal meaning construction.

In addition to the difficulties involved in combining the stages of metafunction analysis, the second possible weakness of the approach is that not all images seem to lend themselves to metafunction analysis in a way that would produce constructive insights. The systemic functional approach treats images and verbal language as analytically corresponding structures. If we adopt this starting point in research, we should be able to assume that *all* images fulfil the metafunctions Halliday proposed for verbal language and can therefore be analyzed by comparing them to predefined verbal structures (cf. Kress and van Leeuwen 2001, 2). Van Meerbergen’s and Chueasuai’s studies both reflect that the analysis method might not be equally applicable to all types of images. Van Meerbergen, for instance, conducted an interpersonal metafunction analysis on the images that display human or humanlike subjects, interacting and performing tasks, and simply skipped the images that displayed something else (e.g. 2010, 136, 176, 189–190). Chueasuai’s study did not include interpersonal metafunction analysis on either of the modes.

All in all, I propose that the images used as examples in metafunction analyses generally tend to represent misleadingly homogenous image types. I here refer to images that display human or humanlike subjects. When examining images like these, it is often straightforward to describe the image participants, their mutual relationships, and their relationship with the viewer of the image. This image analysis procedure can be contested, for instance, by considering how we might describe the metafunctions of an abstract painting or other types of non-representational images that often tend to be beyond attempts of objective verbalization. As Forceville (1999,

172) argues, the “grammar” of images is perhaps not as universal as Kress and van Leeuwen propose. Many technical images, such as the images in my data, also contest the systemic functional way of conceptualizing visual information. I now aim to demonstrate this by discussing one of the images used in the source text employed in my study.



**Figure 2.** The first image of the source text used in the translation assignment of the study.

Figure 2 presents the first image of the multimodal source text used in my study (the entire source text is presented in Appendix 1). A metafunction analysis of this image would start on the ideational level by identifying the participants in it. As the object depicted in the image does not have eyes, the analysis of the interpersonal metafunction would be limited to reporting the perspective and the size of the image frame. Image analysis on the compositional level would refer to listing the color, the positioning and the relative size of the participants. These details can obviously be listed, but one may ask if this would significantly advance our understanding of the image. Further, one can ask if image analysis tools such as the metafunction analysis really denotes a procedure of image analysis, or merely a procedure of *image*

*transcription*, referring to an attempt to bring images into verbal form in order to compare them with verbal language (see e.g. Flewitt et al. 2009, 45).

So far, I have questioned whether the systemic functional image analysis procedure is applicable to all types of images. Another point of critique to the approaches that analyze images by comparing them to the structures of verbal language is that images may have characteristics and properties that do not correspond to these structures. With linguistically-grounded analysis methods, these characteristics may go unnoticed. As Hartmut Stöckl, a prominent scholar in the field of word–image interaction research, states:

The danger of contrasting two modes, however, is that we tend to somehow look at one mode in terms of another. So, mostly, due to language's dominance, we seem to be asking which linguistic properties images have. Thus we run the risk of overlooking some important design features of images which are outside the linguistic perspective. (Stöckl 2004, 18)

Most of the praise received by linguistically-inspired approaches to multimodality evolve around the notion of linguistics possessing finely articulated analytical tools (e.g. Bateman and Schmidt 2012, 32; Bateman 2014, 186) which stem out of a long tradition of research (e.g. Hiippala 2013, 2). I find this justification somewhat unfounded. If I were to exaggerate a bit, I might propose that this equals saying that a kitchen gadget is well suited for car reparation purposes simply because its design is the result of long years of theoretical and empirical probing. Linguistic theories might be a logical and convenient starting point for multimodal analyses, but the use of these analysis tools should be reconsidered if they do not seem to produce useful insights.

One of the guiding notions and starting points of my own research is that we cannot conclusively predict the way in which an image will be interpreted by individual readers by examining the properties of the image only. This idea stems from a growing body of research in visual cognition which suggests that our gaze is directed by our cognitive information-gathering needs to far more degree than it is by the inherent visual properties of the image (for comprehensive reviews on research, see e.g. Mills et al. 2011; Henderson et al. 2007). Viewing images is profoundly task-oriented and directed by our individual preferences, goals and expectations, emotions and prior knowledge (Boeriis and Holsanova 2012, 262). This is why I posit that analyzing images by listing their visual properties provides little insight as to how the images may be interpreted. Different viewers may attribute different meanings to the same image, and the same viewer may attribute different meanings to the same image when examining it for a different purpose.

The same argument can obviously be made about the way readers interpret verbal texts: the idea that each reader interprets verbal text in an individual way has been presented in various lines of research, such as reader response criticism (e.g. Rosenblatt 1978), audience reception theory (Hall 1980), cognitive linguistics (e.g. Langacker 1991), linguistic anthropology (e.g. Ottenheimer 2013), and, naturally, Translation Studies (e.g. Rydning and Lachaud 2010). If we cannot predetermine the way in which words and images will be interpreted in isolation, neither can we predetermine how a reader will interpret their combination. The combination of word and images in an illustrated text is interpreted slightly differently by each reader (e.g. Holsanova 2012, 252). This is the reason I argue that examining multimodal meaning construction from the readers' point of view can offer a welcome addition to the research approaches that are usually adopted in multimodally-oriented (translation) research.

### 2.2.3 Cognitively-Grounded Perspective to Multimodal Meaning Construction

After stating that the approach adopted in my study is *not* systemic functional, I now proceed to describing what it *is*. Coming from outside the commonly employed lines of research, this is not an easy question to answer. For the lack of a more established term, I have described my approach to multimodality as *cognitively-grounded*. By this, I refer to examining the translation of illustrated technical texts against how words and images can be thought to interact based on research into human cognition. The linguistically-grounded approaches to multimodality describe how words and images construct meaning based on linguistic theories. A cognitively-grounded approach describes how words and images construct meaning based on what research into human cognition has suggested about the way in which we interpret the combination of words and images in illustrated technical texts.

The starting point of a cognitively-grounded approach is that we may assume that the readers of illustrated technical texts integrate words and images as they read, and that, for this reason, their interpretation reflects the way in which they interpreted the combination of the modes. The approach entails examining these interpretations. Instead of analyzing the multimodal text per se and dividing it into components and properties, my analytical focus is on the *cognitive effect* of the multimodal text. In other words, my research focuses on the cognitive practices by which we understand multimodal artifacts and the effects of these cognitive practices as displayed in our

personal descriptions of these artifacts (translation diaries) and the renderings of our interpretations of these artifacts (translations).

The theoretical basis of my study, introduced in detail in Article 1, is built on research on the cognitive processes involved in illustrated text comprehension as well as research on translation as cognitive activity. Both lines of research represent the so-called information processing approach of human cognition, and the point of contact between them, in my analysis, is the *mental representation* constructed of multimodal source text information. What we translate is indeed not source text information per se, but mental representations constructed of this information (e.g. Jensen 2010, 216–217). Studies into illustrated text comprehension have proposed two prominent, competing models which offer an interpretation for how mental representations are created in illustrated technical text comprehension, namely the Cognitive Theory of Multimedia Learning (CTML; e.g. Mayer 2002, 2005) and the Integrated Theory of Text and Picture Comprehension (ITPC; e.g. Schnotz and Bannert 2003; Schnotz and Kürschner 2008). My research compares these two models and considers how a translator's comprehension of an illustrated source text would unfold in the light of these models.

CTML proposes that, when a person reads an illustrated technical text, the cognitive system of the reader makes referential links between verbal and visual information and combines these into a single, integrated mental representation (Mayer 2005, 40). ITPC, in turn, proposes that words and images, being fundamentally different forms of representation, cannot be integrated into a single mental representation. Instead, ITPC proposes that the reader of an illustrated text constructs multiple complementary representations for both words and images. The model proposes that when we read verbal text, we construct both a verbal model of the text contents as well as a visual model, representing the typical visual features of what is described in the verbal text. Similarly, the model proposes that when we examine an image, we construct both a visual model as well as a verbal model representing what we see in a verbal form (Schnotz and Bannert 2003, 145–146). ITPC argues that these models cannot be integrated into a single representation – instead, they interact by elaborating and contesting each other (Schnotz and Bannert 2003, 147; Schnotz and Kürschner 2008, 180) (see Article 1 for a more thorough discussion).

If we employ the cognitive models of illustrated text comprehension to the analysis of translation, we can assess what the translator's input might consist of when translating an illustrated text. Following CTML, the input for translation would be a single mental representation consisting of both verbal and visual information.

Following ITPC, the input for translation would consist of various representations – some verbal, some visual – which may both complement and contest each other. What I propose in this dissertation is that regardless of which one of the models is deemed more accurate, the implications these ideas have for Translation Studies remain essentially the same: when translating an illustrated technical text, the translator's interpretation of the verbal text – and, consequently, the translation solution – can be shaped by visual information. Further, if we find a way to scrutinize this interpretation, we may examine it to make inferences about how multimodal meaning construction took place for this particular reader. This is the theoretical position I set out to test in the empirical part of my study.

To sum up, based on a cognitively-grounded approach to multimodality, a translator's interpretation of a multimodal artifact may reflect the way in which the translator constructed meaning from the combination of the modes. Subsequently, this interpretation can be analyzed to make inferences about multimodal meaning construction. In this dissertation, analyzing someone's interpretation refers to examining the translation diaries and the translations, as these are taken to represent research participants' personal descriptions of their interpretations, and renderings of their interpretations, respectively.

The methods I set out to employ and develop in this dissertation are not novel methods for multimodal text analysis. In fact, both of my sources of data as such, the translation diaries and the translations, consist of verbal information only. Instead of scrutinizing a multimodal text per se, this dissertation analyzes multimodal meaning construction by examining data that reflects various people's interpretation of a multimodal text. I propose that this type of analysis may inform our understanding of multimodal meaning construction in ways that analyses based on multimodal texts on their own cannot. The major benefit of the approach is that it enables the comparison of the different ways in which words and images may combine to make meaning from the perspective of individual readers. For instance, Jana Holsanova (2012, 252), a prominent researcher in the field of Cognitive Science, has pointed out that much of the research examining word–image interaction does not take into account that different people interpret the combination of word and image in different ways and hence concludes that it is crucial to include this perspective in multimodally-oriented research. This dissertation is an attempt to test one possible way of adopting such a perspective.

## 2.3 Translation Process Research

As described by Birgitta Englund Dimitrova (2010, 406), translation process research examines “the nature of the cognitive processes involved in translating, with a focus on the individual translator.” Within recent years, translation process research has developed into one of the most active fields of research within the discipline, incited by the development research tools that may be used to produce data from which cognitive processes can be inferred (Hansen 2013, 88). My dissertation, too, examines aspects related to the translation process. However, my research design differs from the mainstream of current process-oriented translation research in both data collection and analysis methods. In this section, I reflect on these differences.

The translation process, as defined by Gyde Hansen (2003, 26), entails everything that happens between the moment when a translator starts reading a source text and the moment when the translator finishes the target text, including reading, decision making, problem solving, referring to sources, writing and editing. Common topics in translation process research include translation competence and its acquisition, metacognition (what translators know about their processes), the impact of technology on translation and the cognitive rhythm of translators, including the analysis of pauses (Saldanha and O’Brien 2013, 111–112). Some studies have also examined the role of revisions and the effect of time pressure on translation (see Englund Dimitrova 2010, 409–410).

Current translation process research typically triangulates the data collected with various data elicitation methods. The data collection tools include keystroke logging software, screen recording, and eye tracking (Hansen 2013, 88). One of the common challenges of this type of research is that each of these tools generates substantial volumes of data, the synchronization and analysis of which can be laborious. This problem can be alleviated by resorting to an asset management system (e.g. Göpferich 2010a), which is a database containing the streams of different types of data. The actual translations produced by research participants are only a part of this data stream. The present decade has hardly witnessed translation process-oriented research that would place as much explicit attention on translations as this dissertation does.

Collecting translation diary data is an example of an *introspection method* in translation process research, referring to asking translators what they thought about during translation (e.g. Hansen 2013, 89). Introspection methods may be divided into two types: introspection taking place concurrently and retrospectively. Translation diaries belong to the latter group. Concurrent introspection methods, or

so-called think-aloud protocols, in which translators intend to verbalize their thought processes while translating, have generally outweighed retrospection methods in translation research. Retrospection methods are generally considered less reliable because they are limited by the constraints of working memory: translators are not likely to remember everything they thought about during translation if the reflection is done after the task is finished (e.g. Angelone, Ehrensberger-Dow and Massey 2016, 45). The possibly non-comprehensive nature of translation diaries as research diaries is discussed in Section 3.5. However, concurrent introspection methods, too, have limitations. For many translators, talking during translation might not feel natural (cf. Hansen 2013, 90) and the unnaturalness of the data collection situation may affect, for instance, the participants' lexical decision-making (Jääskeläinen 2000, 80).

As I discuss below in Section 3.5, the translation diary procedure followed at the University of Tampere at the time of the data collection is highly similar to the Integrated Problem and Decision Reporting (IPDR) procedure introduced by Daniel Gile (2004). Previous translation process research examining IPDR reports written by translation students have set out to answer research questions that differ from mine to great extent. Therefore, these studies have employed analysis methods that are largely different from the method employed in my analysis.

Previous studies examining IPDR reports have primarily aimed to find out how many times a particular aspect of the process is mentioned by the research participants in the data. Mina Baradaran Hezaveh and Gholam-Reza Abbasian (2015) examined the work of translation students, their goal being to identify the types of translation problems the students encountered. The students' reports were analyzed by pinpointing the problems discussed by the students, classifying these into three categories (lexical, structural and pragmatic problems) based on the analysts' empirical observations, and counting how many times each problem type was mentioned in the data. Gyde Hansen (2006) analyzed a group of translation students and compared IPDR with other retrospection methods. The study set out to examine how many problems and decisions were reported by the students when employing each method of data collection, the overall aim being to elucidate the amount of information each method yields. The analysis included counting the problems and decisions described in the students' reports, and comparing these numbers for each data collection method. My own analysis does not focus on how many times a particular aspect of the process is mentioned by research participants, but *how* a particular aspect (namely word–image interaction) is *described* by

participants. Such a research question naturally requires an analysis method that focuses on the nuances of the participants' descriptions.

# 3 DATA COLLECTION

The data of my study comprises two parts: the translations of an illustrated source text, produced by the participants, as well as the translation diaries the participants wrote during the translation assignment. This chapter reports how the data was produced. Section 3.1 introduces how the source text was created for the translation assignment. Section 3.2 explains how the data was collected from the participants and Section 3.3 discusses a group meeting that was held with the participants after they had handed in their translation assignments. In Section 3.4, I discuss the research participants and reflect on the implications of examining student translators as opposed to experienced professionals. Finally, in Section 3.5, I reflect on the translation diary procedure followed at the University of Tampere as well as the general characteristics of the translation diaries in my research data.

## 3.1 Design of the Source Text

The source text used for the translation assignment (Appendix 1) presents the illustrated operating principles of two different types of wet magnetic separation machines, concurrent drum separators and counter-current drum separators. The devices are used in the mining industry for ore beneficiation and their working principle is based on the induction of magnetism. The source text, the verbal text as well as the images, was produced with the assistance of the staff of the Eastern Finland office of the Geological Survey of Finland (GSF). The text was written in English and it was proofread by Dr. Richard Browner from the Western Australian School of Mines. The source text consists of just over 500 words and two large colored images. It is represented as a chapter of a basic mining engineering text book introducing different ore beneficiation methods. The primary target audience of such a text would typically be students of mining engineering. As confirmed by GSF staff members, the text represents the particular text type well for both its content and style. The illustrations are generalizations of the operating principles of the devices and do not represent a particular commercial model.

The subject area for the source text was selected based on two criteria, both related to maximizing the students' involvement with the images. First, I assumed that mining technology might be a subject domain in which the students would have poor background knowledge – research on illustrated text comprehension has shown that the less the readers have prior knowledge of the subject domain, the more they resort to the images for their information gathering needs (Mayer and Gallini 1990, 724; Hegarty and Just 1993, 736). My second criterion was to select a subject area in which established terminology might be scarcely available in Finnish: if the required terminology had been readily available, it would have been less likely for the students to resort to the images when coming up with translation solutions. Both of these assumptions were later confirmed by the students, as discussed below.

The images and the corresponding descriptions of verbal text are presented in the source text in close proximity to each other. Research into illustrated text comprehension has established, perhaps quite logically, that in order to promote the comprehension of illustrated texts, both modes should be made available for inspection simultaneously so that readers can make frequent switches between the two. Physical continuity in space reduces some of the cognitive load associated with cognitively integrating the two modes (Hegarty and Just 1993, 738; Mayer 2002, 344–345). The students received both pages of the source text printed on one A3 size sheet. The reason for the large size of the print was to make sure that both of the images of the source text would be available to be viewed concurrently with the verbal text. Had the source text been printed on two separate papers of size A4, for instance, a part of the verbal text explaining the operating principle of the counter-current magnetic separator would not have been on the same page with the image displaying the operating principle in a visual form.

The main reason for my decision to produce the source text myself as opposed to using an existing illustrated technical text was that I wanted the source text to include diverse examples of word–image relationships: I wanted to include instances in which the verbal and the visual would also contest each other. The logic behind this idea is that if the information conveyed by words and images had been *symmetrical* – in other words, expressing more or less corresponding information – it would have been impossible for me to deduct whether the translation solutions were based on verbal or visual information, or both. I hence decided to modify the relationship between words and images in certain parts of the source text so that the information conveyed by the two modes was, in one way or another, *asymmetrical* (for a thorough account of the modifications made, please see Section 2 of Article 3).

For instance, in one section of the source text, visual information was deleted from the image: the verbal text accurately described a particular part of the operating process (tailings or nonmagnetic particles exiting the separator) but the corresponding information was not presented in the image. On two occasions, the visual and verbal texts were modified so that the information provided by the two was straightforward contradictory (the verbal text expressed that the device was submerged under water, while only the bottommost part of the device was under water in the image, and the verbal text expressed that a certain part of the device was located in the upper part of the device while, according to the image, the part was in the lower part of the device). Further, the shape of a part of the device called launder was modified: while the term typically refers to a trough or a long, narrow container, it was presented in the image as nearly square-shaped. The rationale behind these modifications was that the asymmetry of information might make it easier to distinguish which mode the translation student considered to be of more relevance during translation.

## 3.2 Implementation of the Translation Assignment

The data of the study was produced at the University of Tampere by a group of Master's level English translation major students as a part of a Technical Translation Seminar from English into Finnish, taught by Dr. Hilkka Pekkanen, in the spring of 2014. The number of students taking part in the seminar was eight. While the translation was a compulsory assignment for the course, submitting one's work for research purposes was voluntary. All eight students gave permission to have their translations and diaries analyzed in the study. The research project was submitted for evaluation by the university's ethics committee for research involving human subjects and it was exempted from the need for a review.

The translation was conducted as a homework assignment, and the students had one week to complete it. The decision to conduct the translation as a homework assignment as opposed to an in-class assignment came down to allowing the students more time to write the translation diaries: this way the students had the opportunity to write their diaries in a more relaxed schedule and therefore, perhaps, in more detail. Since it was in the best interest of the study that the students treated visual information exactly as they normally would, they were not informed of the research questions prior to conducting the translation assignment. This also meant that the students were not specifically instructed to comment on the images in their

translation diaries. This set-up made it possible to determine whether the students themselves decided to comment on the images as an intrinsic part of the source text. The students were told that the source text was an extract of a mining engineering text book and that the target audience would be readers interested in ore beneficiation, such as engineering students.

At the time of the data collection, translation students at the University of Tampere were expected to analyze their translation practice by writing a translation diary for almost all of the translation assignments they were given during their studies, starting from the first courses on the Bachelor's level. One could therefore propose that by the time they had reached their Master's level studies, the students were well acquainted with the process of writing the diaries and that the data collection method did not have a negative impact on the ecological validity of the setting. The diaries were written in Finnish, in keeping with the instructions provided for the particular translation course. My analysis of the diaries was conducted on data in Finnish, and in Article 2, excerpts of the diaries were translated into English by me.

The students were allowed to discuss the assignment among themselves during the one-week period. There were two reasons for this. First, it would have been simply impossible to control that they would *not* do so – the students would meet each other regularly on other classes, and some of them might be friends and see each other during their free time. Second, I believe this encourages a collegial spirit among translation students and develops productive practices for future working life. When asked about this during the post-translation meeting, some of the students mentioned having discussed how to translate the names of the devices into Finnish – the established terms in the specialized domain are somewhat monstrous *myötävirtamärkärumpumagneettierotin* and *vastavirtamärkärumpumagneettierotin*, so these were indeed challenging parts of the assignment. No one mentioned discussing issues related to the images.

The students were also given a short questionnaire (see Appendix 2) to fill out, partly before, partly after the translation assignment. In the questionnaire, the students were asked to state their native language and then to assess, on a scale from 1–5, their background knowledge of the subject matter and how well they felt they understood the subject matter after translation. The questionnaires confirmed my assumption of the students having poor background knowledge in the chosen subject domain. Four of the students assessed their background knowledge as very poor, four as poor (options 1 and 2 on the questionnaire). Even though these estimations are rather subjective, the fact that none of the students reported being

considerably better acquainted with the subject matter than the rest of the group adds to the comparability of the data. At the back of the questionnaire, there was another copy of the source text, on which the students were asked to mark the sections they found particularly challenging to translate. My original plan was to include the replies for this last question in the analysis of the data. However, I noticed that the students discussed exactly the same challenges in their translation diaries as well – perhaps prompted by this question. To sum up, even though I initially planned to incorporate the questionnaires to the analysis of the data, in the end, they were only used to evaluate the subject domain related background knowledge of the participants.

### 3.3 Post-Translation Group Meeting

I met with the students a week after completing the assignment. This meeting was principally meant to serve two purposes. First, as the translation assignment also had a pedagogical purpose in the course curriculum, the meeting included giving the students feedback on their translations and addressing their doubts about the source text and the subject matter. The second purpose of the meeting was to inform the group about the aims of the research project and to ask for permission to use their work as data in the study. In other words, the main purpose of meeting with the students was to give them feedback and inform them about the study in which they participated. However, I asked for the students' permission to record the conversation as I suspected that the students might provide some additional insight as to why they made certain solutions, which indeed turned out to be the case.

In two of the published articles of this dissertation (Articles 2 and 3), I have referred to this meeting as a *group interview* since the discussion was recorded, transcribed and used to inform the analysis, but the meeting had several characteristics that do not entirely fit what is commonly understood by a group interview in research. Roller and Lavrakas (2015, 104–5) describe group interviews, or group discussions, as collecting research data by gathering a group of participants and encouraging them to talk about their divergent and convergent thoughts or ideas on a certain topic. The role of the moderator of a group interview is typically considered to include keeping the discussion on topic, following a more or less structured set of predetermined questions, while encouraging everyone in the group to interact freely (Morgan 2011, 146; Roller and Lavrakas 2015, 104–7).

In the group meeting of my research project, the focus of the questions asked and the topics discussed was not to elicit more research data for my own purposes,

but to provide the students with an opportunity to reflect on their own work, to ask others what they had done and to express any doubts they had been left with after completing the assignment. If regarded as a data collection session, the main shortcomings of this meeting was, perhaps, that I did not aim to inclusively and comprehensively enquire about *everyone's* views and experiences. Some more active students dominated the discussion slightly and the quieter ones were not prompted to speak more, even though this can never be fully avoided in group discussion situations (e.g. Barbour 2007, 82). The meeting lasted an hour and a half, which is the length of a standard class at the university.

In the beginning of the meeting, I told the students the aim of my study and explained how the source text had been produced. The students appeared to be surprised to hear that the focus of the study lied in the role of images; none reported having considered this. We then discussed the physical basis of the operating principle of the devices, and proceeded to analyze the translations in a somewhat linear manner. The focus of the discussion tended to be on comparing the terminological solutions the students had used. When asked about the sources they had consulted during translation, the students confirmed that it had been difficult to find relevant terminology from parallel texts written in Finnish. As described above, this observation is connected to my subject matter selection criteria.

There were three major insights gained during this meeting that I used to inform my analysis of the translation diaries in the conclusion section of Article 2. First, when discussing the images and the instances in which they seemed to contest or contradict the verbal text, one of the students reported not having noticed these instances at all. The student concluded having looked at the images in great detail before reading the verbal source text, but then having disregarded the images while performing the actual translation. The other seven affirmed having looked at the images, which was also confirmed in their translation diaries, as all seven discussed issues related to visual information. Second, several students discussed the part of the source text in which visual information had been deleted and mentioned that it had not occurred to them that the image could be faulty in any way. They had simply thought they were “too stupid” to understand the process correctly. Third, one of the students reported having thought that the fact that a part of the process was missing from the second image of the source text could well have been due to errors produced when printing the source text. This, too, reflect a view according to which images are seldom faulty – the error was attributed to the reproduction rather than the production stage of the image.

### 3.4 Research Participants

The research participants were a group of eight translation students, four females and four males. The translation course during which the data was gathered was called *Technical Translation Seminar*, which was a part of a Master's program concentrating on specialized translation. All of the students had received a Bachelor's degree in English translation at the University of Tampere, having therefore completed both theoretical as well as practical translation courses (both from and into English). All of the research participants spoke Finnish as their native language. They therefore wrote the translations in the language they have the highest command of, and inadequate language skills as a factor affecting the translation outcome could be ruled out. An important question to ask about the research setting of my study is how the results would have been different had the research participants been professional translators. In this section, I will reflect on this question by reviewing studies that have set out to compare the performances of student and professional translators. I will then return to this question in Section 7.2 – in other words, after reviewing my research findings – to assess how these findings may have been affected by the student status of the participants.

The comparison of student and professional translators is often done by reflecting on the quality of translations produced by these groups. Unfortunately, the multifaceted concept of translation quality is not explicitly defined in the studies quoted below. Yet, as they mainly set out to evaluate the mistakes made in the translations, one might propose that they foster a view in which high translation quality equals an error-free product (cf. Conde 2008, 50). Even though some studies have found no apparent difference in the quality of translations produced between the two groups (see e.g. Kiraly 1995, 107; Séguinot 2000, 143), the most commonly made observation – perhaps unsurprisingly – is that professional translators produce higher quality translations than students.

Susanne Göpferich's (2010b) study, for instance, compared the translations of a technical text made by a group of 12 translation students and a group of 10 professional translators, all with at least ten years of experience. The study concluded that the professional translators produced higher quality translations than the students. The professionals translated faster, made fewer mistakes, and proceeded with the assignment in a more strategic manner than the students did. Yet, Göpferich (2010b, 49) reports having been negatively surprised with the performance of the professional translators, stating their translations were “far from what would have been expected of a translation expert.” One of the possible explanations Göpferich

offers for this is that the subject matter of the source texts used in the study might not have represented the domains of expertise of the professional translators (see also Jääskeläinen 2010, 219). No translator, even one with ample experience in technical translation, will be familiar with all possible subject matters, and coming to terms with particular subject matter-specific terminology for the first time is challenging even for experienced translators.

One could argue that the observations I make in the study would have more value if the research participants had been experienced professional translators. Yet, recruiting a group of experienced technical translators might not have offered guarantees of receiving translations representative of translation expertise, as Göpferich puts it (2010b, 49). Further, recruiting a group of technical translators with particular domain-specific expertise might have been exceedingly difficult: as mentioned in the Introduction, the staff of the Geological Survey of Finland affirmed that translators with ore beneficiation-related background knowledge are very hard to come by. Moreover, conducting the study with such translators might not even represent the reality of the translation industry – it might well be that most ore beneficiation-related commissions in Finland will be completed by a translator who is new to the particular subject domain.

There have also been other arguments in favor of using student translators instead of professionals. Hans P. Krings (2001, 72) has proposed that using professional translators in empirical research settings might be counterproductive because professionals are easily disturbed by having someone observe their weaknesses and shortcomings. For translation students, this is commonplace. Further, as Juan Sager (1994, 152) emphasizes, the most important variable in the translation process is the *diversity* of human translators, and translation students represent this diversity well with their varying levels of background knowledge. A further advantage is that translation students at the university in which the study was conducted were accustomed to validating their translation choices carefully in their translation commentaries. Translation students might therefore be more likely to produce in-depth reports of their translation process than experienced translators who might not have been asked to comment on their thought processes in a written form in years.

### 3.5 Translation Diaries as Research Data

The translation diaries produced by the participants could be described as semistructured: instructions for writing the diaries had been provided in the first year of the study program, but had not been regularly reinforced. The students had been instructed to describe the communicative situation (commissioner of the translation, the target audience, and so on) and the source text (style, text type, and subject area), to describe the translation process (macro and micro level strategies), to discuss the problems they encountered during the translation process, the steps they took to solve them, and the rationale behind their decisions, as well as to introduce the references that were consulted during the translation assignment (UTA käänöskomenttiohje).

In the data of the study, these general guidelines were followed to a reasonable extent. The students concentrated mainly on discussing the features of the target text and describing their translation process. A few, but not all, discussed the communicative situation and reflected upon their own work. The most widely discussed topic was the terminology. Six out of eight students mentioned terminology as the most difficult feature of the target text. The terminological difficulties were mainly reported to be due to the poor availability parallel texts. Given that domain-specific terminology is one of the most distinguishable features of technical texts (e.g. Byrne 2012, 51; Olohan 2016, 26), it is perhaps not surprising that the discussion of microstrategies concentrated on mainly on terminology: all students introduced individual terms they found particularly challenging to translate.

The translation diary procedure followed at the University of Tampere at the time of the data collection is highly similar to the Integrated Problem and Decision Reporting (IPDR) procedure introduced by Daniel Gile (2004). IPDR sets “a systematic requirement for written introspective reporting by students whenever they hand in a translation assignment” (Gile 2004, 15). The semistructured reports are collected in a written form and include discussion on the problems the students encountered during the translation process, the steps they took to solve them, and the rationale behind their decisions, as well as the sources and references that were consulted during the task (Gile, 2004, 3–4).

As Gile affirms, this type of reporting during translation benefits the students and the instructors alike. From the perspective of the student, it increases an awareness of the translation process and emphasizes that translation is a demanding operation requiring intense decision-making. From the perspective of translation instructors,

the procedure elucidates what the students are doing and what they find particularly difficult, and supports the instructor in identifying translation strategies (2004, 4–9).

Gile (2004, 8–9) also reflects on the potential of introspective reports as data for research, emphasizing that while IPDR does not offer a revolutionary way of accessing information that is not available through other methods, the method does have its benefits compared to other introspection methods, such as think-aloud methods. One obvious benefit is that the data is easily gathered and available to the researcher in a directly readable form. Moreover, this type of introspective reporting does not include distraction between translating and verbalizing one's thoughts at the same time (cf. think-aloud methods), nor does it require the students to work under strict time limitations or with a particular software (*ibid.*).

The flexible, semistructured nature of these introspective reports can be considered an advantage for the students as it allows for creativity in the writing process. Yet, when examined as research data, this flexibility can also result in the reports being non-comprehensive: the researcher does not have control over what the research participant discusses in the diary, nor does the researcher have an opportunity to encourage the research participant to reflect on matters further or to ask clarifying questions. Gile (2004, 10) suggests that this shortcoming could be remedied with using a more structured set of questions. Indeed, both Gile (2004, 10) as well as Göpferich and Jääskeläinen (2009, 172) conclude that the main limitation of IPDR and similar forms of semistructured introspection as methods of data collection is the unpredictable nature of the data: the contents of the introspection reports depend entirely on what the participants themselves regard as relevant.

However, this very notion justifies my selection of data collection method. The reason I did not inform the students of the research questions of the study – and, consequently, did not instruct them to comment on the images in their diaries – was to find out if the students indeed regard visual information as relevant enough to analyze as a part of their translation process. In addition, as discussed towards in Section 4.1, translation diaries written in response to a structured set of questions could not be plausibly used as data in a phenomenographic analysis. Yet, the partly non-comprehensive nature of the diaries as research data did become evident in the post-translation group meeting. As discussed above, during the meeting, some of the students offered some interesting insights into their translation process and these insights supported the rest of my analysis. In my research, it became evident that while semi-structured translation diaries proved to be a highly interesting source of data, obtaining a thorough understanding of the reasons behind the translators' solutions requires a complementary data collection method.

## 4 ANALYSIS METHODS IN THE EMPIRICAL ARTICLES

This chapter introduces the analysis methods I employed in the empirical parts of my dissertation, namely Articles 2 and 3. As Andrew Chesterman (2007, 1) writes, “methods are the ways in which one actually uses, develops, applies and tests a theory in order to reach the understanding it offers.” The methods chosen for this particular study aim to test and develop the ideas put forward in Article 1, the theoretical basis of the dissertation: based on the cognitive models of illustrated technical text comprehension, we may assume that a translator’s interpretation of the verbal text – and, consequently, the translation solution – may be shaped by visual information. My empirical methods aim to test and develop this argument, and elucidate how the multimodal meaning construction might unfold in practice. Section 4.1 introduces phenomenographic analysis, employed in the analysis of the translation diaries, and Section 4.2 introduces Choice Network Analysis, employed in the analysis of the translations. In these sections, I reflect on the strengths as well as the inherent biases of each method and discuss how I have chosen to interpret the results of my analyses.

### 4.1 Phenomenographic Analysis of the Translation Diaries

This section introduces my analysis of the translation diaries, which aimed to elucidate *if* and *how* the students commented on the role of images during the translation. The first question was answered by reading through the data: seven out of eight students discussed images and word–image interaction in their diaries. The research approach adopted for answering the second question was phenomenography. Phenomenography sets out to map the different ways in which various aspects of the world are experienced, conceptualized, perceived, and understood by different people (Marton 1994, 4424; 1988, 144). The approach is hence interested in the *variation* of ways in which people experience and

conceptualize things. The emphasis on variation is precisely why the approach was adopted in the present study. My phenomenographic analysis aims to elucidate the diverse ways in which the participants conceptualize word–image interaction.

I start this section by discussing the conceptual basis of phenomenographic enquiry. In Section 4.1.2, I address the critique that has been expressed towards phenomenography. Finally, in Section 4.1.3, I introduce a detailed description of the procedure I followed in my analysis, as this is often regarded as a vital component in assessing the reliability of phenomenographic enquiry.

#### 4.1.1 Phenomenographic Inquiry

Phenomenography sets out to map the different ways in which various aspects of the world are experienced by different people (Marton 1988, 144; 1994, 4424). The research approach originated at the Department of Education at the University of Gothenburg, where research on the experience of learning suggested that a group of students reading the same text understood it in qualitatively different ways (Marton 1988, 148–152). Even though the approach was developed and has mainly been applied in education research, its aim transcends the educational context: phenomenography sets out to identify similarities and differences in the way we experience and comprehend phenomena in the world around us (Marton 1994, 4429). Phenomenographic inquiry hence concentrates on the *variation* in the ways of comprehending things. Just as there is probably always variation in the ways students comprehend subject matter presented to them in a textbook (Marton 1981, 184), there is probably always variation in the ways a group of translators comprehend the same source text.

Marton describes phenomenography as a means to examine “the different ways in which people *experience, interpret, understand, apprehend, perceive or conceptualize* various aspects of reality” (1981, 178; emphasis added). This description invites us to think that these verbs are to be handled as synonyms or near-synonyms. We hence understand experiencing as understanding and interpreting something – assigning meaning to phenomena as they are encountered. Phenomenography should be understood as an integral perspective on research rather than a single method of analysis. It is underpinned by a nondualistic ontological stance: the subject and object of an experience are not considered as separate entities (Säljö 1997, 173). Experience or conception is understood as the internal relationship between the experiencer and the experienced (Marton and Booth 1997, 121). Phenomenography is characterized

by the adoption of a so-called second-order perspective: instead of making statements about how things “really are” in the world, the approach aims to make statements about how things appear to people (Marton 1981, 178). Research conducted from the first-order perspective might ask questions such as “How do words and images interact in illustrated texts?” Within the second-order perspective adopted in this part of my study, however, I ask “How do participants themselves experience and conceptualize word-image interaction in illustrated texts?”

Phenomenography should be clearly distinguished from *phenomenology*, founded by German mathematician and social thinker Edmund Husserl as a theoretical philosophical movement in the early years of the 20th century (Zahavi, 2003) and developed by various philosophers such as Heidegger (1962) and Merleau-Ponty (1962). Even though the object of research in both phenomenography and phenomenology is considered to be the human experience, there are fundamental differences between the two. Whereas phenomenology adopts a first-order perspective, describing phenomena “as they are”, or the *essence* of phenomena, the second-order perspective adopted in phenomenography aims to describe phenomena as they are understood (Barnard, McCosker and Gerber 1999, 213–214). In other words, phenomenology involves the study of reality as it appears to individuals; phenomenography focuses on studying individuals’ conceptualizations of reality (e.g. Gall, Gall and Borg 2007, 491). However, as Cibangu and Hepworth (2016, 148) remark, a thorough comparison of phenomenology and phenomenography is complicated by the fact that the former draws on over a century of rich literature, with various different branches that differ in their aims.

The data used in phenomenographic analyses is collected from a group of people individually reflecting on their experience of a phenomenon. The most commonly adopted method of data collection in phenomenographic research is the individual interview. However, some studies have also employed group interviews, written responses and even children’s drawings as research data (Marton 1994, 4427). There are also examples of using unstructured learning diaries as research data (Prinsloo, Slade and Galpin, 2012). The greatest advantage of the individual interview as the data collection method is considered to be the ability of the interviewer to encourage the interviewees to maximum reflection by using follow-up questions. Nonetheless, as discussed in the following section, collecting data by means of interviews is also one of the most commonly discussed methodological limitations of phenomenography.

The data of my own phenomenographic analysis was not produced in response to series of questions formulated by the researcher and the study may hence be

connected to what Biörn Hasselgren and Dennis Beach (1997, 197) term *naturalistic phenomenography*. The idea is built on Yvonna Lincoln and Egon Guba's (1985) description of the naturalistic method of inquiry, introduced as a paradigm in which the researcher avoids manipulating research outcomes. Naturalistic phenomenography entails collecting data that is not produced inclusively for the purposes of the research. This refers to recording what is expressed in a given situation "without *direct* manipulation or involvement from the researcher" (Hasselgren and Beach 1997, 197; emphasis added). Examples of naturalistic data collection in other phenomenographic studies include recording classroom discussion in which the researcher does not participate (Lybeck 1981). The acquisition of the translation diaries in this research project could be described as naturalistic: in their diaries, the participants reflected on word–image interaction in ways that they considered important, without the involvement of and without answering questions formulated by the researcher.

In the above quote, the word *direct* is significant: in naturalistic phenomenography, the researcher is not directly involved in the participants' reflections of a topic. Yet, as Peter Kelly (2003) observes, even if there is no direct dialogue between the researcher and the research participants, the researcher may still influence the social context of the participants' reflections. This is also true for my study: the students knew that I would read their diaries and that, with their permission, the diaries would be used for research purposes and this may have affected what they wrote.

The effect of my researcher involvement in the participants' reflections on the role of images in the source text, however, could be said to be decreased by the fact that the students did not know it was indeed word–image interaction I was interested in. It is therefore unlikely that my involvement would have affected the way in which this particular issue was reflected on. The researcher involvement in the present study may also have had a positive effect on the translation diaries produced: knowing that their diaries might be used for research purposes, some of the students may have spent more time on writing the diary than they otherwise would. I refer to this as a positive effect because producing a thorough translation diary also benefits the students themselves.

The actual phenomenographic analysis starts with a thorough reading of the data and selection of expressions (direct quotes) that refer to experiencing the phenomenon under study in a particular way. These quotes are arranged into groups referred to as *categories of description* based on their similarities and differences (Marton 1988, 145). In other words, the categories are not selected in advance, but emerge

from the data as quotes are compared. The same individual may express more than one way of conceptualizing a phenomenon. For this reason, the individual – in the case of my research, the individual translation student – is not the unit of analysis (Marton and Pong 2005, 346). Instead, phenomenographic analysis aims to separate “forms of thought both from the thinking and the thinker” (Marton 1981, 196).

The categories of description need to be meaningfully related to each other; after all, they are the result of people reflecting on the same phenomenon. As some ways of experiencing a phenomenon may be more comprehensive than others in relation to a particular criterion, it is often possible to establish a hierarchy between the categories of description, in which some categories subsume others (Marton 1994, 4426). Phenomenographic analysis aims to identify the relationships or the structure between the categories. The main outcome of phenomenographic research is the structured set of categories referred to as the *outcome space* of the phenomenon in question (Marton 1994, 4424). The outcome space should not be claimed to be exhaustive, since the average number of research participants tends to be relatively small – this study being no exception. The analysis merely aims to be *complete* in the sense that nothing has been left out of the outcome space reflecting the collective experience of a particular group as reflected in the data of the study (Marton and Booth 1997, 125).

#### 4.1.2 Evaluation of the Research Approach

The critique posed to phenomenographic research generally revolves around four issues. The first regards the interpretive neutrality of the phenomenographic researcher. The second concerns the challenges involved in collecting phenomenographic data by interviews. The third is related to whether the research participants' verbal accounts really reflect their experiences of the subject(s) of the research. Finally, the fourth is related to the fact that clear instructions cannot be given for carrying out a phenomenographic analysis. This fourth issue will be discussed in the following section in which I discuss my analysis procedure. This section, then, is dedicated to reflecting on the first three.

The researcher's ability to provide neutral interpretations and observations of the data is often brought up when assessing phenomenography as a research method (e.g. Webb 1997, 201; Alsop and Tompsett 2006, 244). This discussion is, by no means, particular to phenomenographic analyses, but characterizes all qualitative inquiry. Eleanor Walsh reflects on whether phenomenographic analysis really

involves “looking into the transcripts to discover the particular ways in which people understand the phenomenon,” or whether it rather reflects the researcher “*consciously* interpreting the data, choosing and discarding data, and thereby *constructing* the relationship” (2000, 20; emphasis added).

Given that data analysis in qualitative research is always an interpretative process and that it is the widely accepted that an interpretative process is never fully objective (cf. Åkerlind 2005, 330), I may conclude that the results of my analysis, by necessity, reflect my own conscious construction of the phenomenon. In phenomenographic terms, my research findings reflect my own experience and conceptualization of the data (cf. Marton and Booth 1997, 125). The only way for the phenomenographic researcher to remedy this lack of objectivity is to document the interpretation process well (see following subchapter) and to provide empirical evidence that supports this interpretation. In phenomenographic research reports, the latter is done by presenting direct quotes from the data.

The second theme that frequently emerges in critique to phenomenography has to do with the data collection methodology. The creators of phenomenography warmly recommended the interview as a means for data collection. In his description of the data collection process, Marton (1988, 154) emphasized that the research participants should be allowed to choose “the dimensions of their answers”, since these dimensions reflect what the participants themselves consider relevant. Yet, Marton (1994, 4427) also proposed that the interviewees could be guided to discover understandings of the phenomenon they were not aware of prior to the interview. Marton and Booth (1997, 129–131) even suggested that the interviewer could break down the interviewees’ defense structures and bring them to a state of “meta-awareness.” However, this notion has been criticized, for instance, by John Richardson (1999, 69) who, quite validly, asks how the interviewer could know that the interviewee has indeed been genuinely brought to such a state.

Other studies have concluded that conducting the interview requires great sensitivity from the part of the interviewer. As Alsop and Tompsett (2006, 257) point out, there may be a thin line between teasing out comprehensive replies to the researcher’s questions and guiding the interviewees’ replies to a direction that no longer corresponds to their own conceptions. Ashworth and Lucas (2000, 304) acknowledge that the questions posed during a phenomenographic interview are easily based on the researcher’s own presumptions about the phenomenon or the participant. To sum up, the main difficulty in collecting interview data for the purposes of phenomenographic analysis appears to be finding the balance between

encouraging the participants to maximum reflection and allowing them to decide the dimensions of their reflection.

The challenges involved in collecting phenomenographic data by naturalistic means are different. Naturalistic data collection does not include asking for clarifications or inciting the research participants to discover new aspects of their own thinking and it is therefore possible that the data does not represent extensively comprehensive reflections of the topic. This means that the data and the results of my phenomenographic analysis are not likely to represent *all* of the possible ways in which this particular group of translation students conceptualized word–image interaction in this particular illustrated text. This also became evident in the post-translation group meeting, as I have already discussed. However, as described above, the results of phenomenographic analyses should never be claimed to *exhaustive*, but merely *complete* in regard to the actual data (Marton and Booth 1997, 125). Naturalistic data collection methods also have their benefits when compared to interview methods: the data can be said to more genuinely reflect the participants' own conceptions of the phenomenon being discussed as the researcher is not able to guide the participants' accounts. Having gathered diary data, I can conclude that I did not impose my own, perhaps partly underlying assumptions about the role of images in translation on my research participants in the form of questions.

The third point of critique frequently posed to phenomenographic inquiry is whether the participants' verbal accounts really reflect what they think. Roger Säljö (1997, 177) has pointed out that one of the factors restricting the conclusions that can be drawn from phenomenographic analyses is that the interviewees may have varying motives for saying things in interview situations. They might say something casually without really meaning it, or they might say something they expect the interviewer to want to hear. Similar challenges, naturally, occur in all research that resorts to interview methods in their data collection (e.g. Sulkunen 1990, 264). These factors need to be taken into account in all phenomenographic research. However, I suggest that this issue might be slightly more sensitive in interview situations than it was in my data collection. If compared to an interview situation, I suspect that my research participants' reflection on their own reasoning has been at least slightly more thought-out, after all, they had to write it down, knowing that their reflection would form a compulsory part of a compulsory translation assignment. Further, the fact that my research participants did not know the aim of my study when writing the diaries makes it less likely that the way in which they commented on the role of visual information would have been influenced by a desire to analyze the topic in a way that might please my interests.

### 4.1.3 Procedure of Analysis

As in all qualitative research, the results of a phenomenographic analysis are the researcher's own interpretations. Assessing the reliability of a phenomenographic analysis hence comes down to assessing the reliability of the researcher's interpretation process (Sandberg 1996, 137). This means that I as the researcher have to detail the procedure I have followed and the decisions I have made during the analysis of my data, as well as to demonstrate that the categories of description I have created are indeed supported by empirical data by presenting them with example quotes from the data (Francis 1996, 44; Åkerlind 2005, 331–332). In this section, I introduce the procedure followed in the phenomenographic analysis of my study (the same description is also presented in Article 2 in an abridged form).

As Marton (1988, 154) concludes, exact rules for the procedure of phenomenographic analysis cannot be specified: the analysis "takes some discovery" for which no algorithms can be provided. Alsop and Tompsett (2006, 245) critique this lack of established guidelines as a procedural weakness of phenomenography and suggest, perhaps slightly pejoratively, that the researcher "is expected to gain an understanding of the process from reading a sufficient number of case studies." This is, in fact, how I have arrived at my own procedure of analysis. My procedure is mainly based on some general guidelines provided by Marton (1988, 154–155; 1994, 4428), as well as Larsson and Holmström (2007, 57). However, my understanding of how the process could be carried out is likely to have been influenced by all of the phenomenographic literature I familiarized myself with. Even though I do not directly quote either of these studies, I wish to highlight the doctoral dissertations by Hilkka Roisko (2007) and Markku Ojaniemi (2013) as reading that particularly enlightened my understanding of how phenomenographic analysis may be implemented in practice.

The stages of analysis outlined by Marton (1988, 154–155; 1994, 4428) and Larsson and Holmström (2007, 57) differ in the number of stages introduced and the use of some terminological choices, but the key elements provided are the same. These stages could be roughly outlined as follows:

1. Reading through the whole data set various times.
2. Reading again, marking the passages in the data where the participants commented on the aspects relevant to the research question(s). Extracting the relevant passages, meaning the direct quotes.

3. Grouping the quotes based on their similarities and differences (focus on the differences between the quotes), then establishing the relationships between the groups (focus on the differences between the groups).
4. Establishing a structure in the outcome space.

I now proceed to describe how my analysis was carried out in practice. The first stage of my analysis included reading through the data various times. The data consisted of eight translation diaries ranging from 400 to 900 words. I noticed that seven out of the eight participants had commented on issues relating to visual information and therefore concluded that the analysis was indeed possible. When reading through the data, my aim was first to gain a tentative understanding of what the students say or “sensitizing” myself to their experience (cf. Ashworth and Lucas 2000, 304). Then, with each rereading, I aimed to obtain a more comprehensive understanding of the data as a whole.

In the second stage of the analysis, I began to select parts of the data that answered the question “What do the participants say about images or the interaction of images and words?” This guiding question was deliberately formulated as a rather open one in order to avoid misjudging what the students regarded as relevant. For instance, when reading the students’ reflections about resorting to other images – images they had found online that could additionally help them during translation – my first instinct was to disregard these comments as irrelevant. I selected the quotes, however, and was later able to observe their connection with other comments relating to resorting to visual information whilst translating.

All in all, the quotes selected at this stage ranged from individual phrases to parts of longer reflection. The selection process proceeded slowly and progressively, and some of these decisions also had to be reconsidered in the subsequent stages of analysis. At the end of this stage, I had gathered a collection of quotes, or the “data pool” of my study, consisting of 53 quotes. The translation diary which included no references in regard to the visual was, inevitably, left outside of the analysis at this point.

In the third stage of the analysis, my attention was shifted from the individual participant to the meaning embedded in the data pool as a whole. I moved on to identify the different ways of understanding the phenomenon, the guiding question being “What are the different ways in which the participants conceptualize the interaction of verbal and visual information?” I read through the quotes again various times and begun to perceive some general themes that ran across the pool. The quotes were arranged into (tentative) groups based on their similarities and differences. In practice, this was done five to six times and then refined again in the

later stages of analysis. Eventually, this led to establishing clearer borders between the groups. It was then possible to detect and determine the distinguishing features of each group. These groups now formed the categories of description of my study.

The final stage of the phenomenographic analysis is usually identified as establishing the structure of the outcome space in which all of the categories of description are present (e.g. Larsson and Holmström 2007, 57). Yet, it was my personal impression that completing the previous stage of analysis, namely establishing the distinguishing features of the categories, was impossible without simultaneously considering the relationships of the categories in the overall structure of the outcome space. In fact, Marton (1994, 4428), too, has proposed that the different stages of analysis sometimes need to be considered simultaneously. In other words, in my analysis, the structuring of the outcome space proceeded hand in hand with the previous stage. This outcome space (see Article 2, page 24) constitutes the final result of the analysis: it represents a contrastive comparison of the different ways of conceptualizing the interaction of verbal and visual information in the particular source text by the particular group of translation students.

## 4.2 Choice Network Analysis of the Translations

In the second part of my empirical analysis, I aimed to examine if and how the word–image interaction is reflected in the translations produced by the students. The analysis method adopted for this purpose is Choice Network Analysis, first created by Stuart Campbell (2000). The method compares the translations of the same source text by multiple translators. The method aims to empirically outline the different ways in which the same source text segment was translated and reflect on the possible motives behind the solutions. In my own analysis, I set out to assemble the variety of translation solutions produced for a particular source text segment, and analyze whether these solutions corresponded to verbal or visual information, or a combination of these.

I start this section by introducing the method and the basic rationale that underlies it. In Section 4.2.2, I describe how choice networks have been created in past research and introduce how I created them in my own analysis. Choice Network Analysis has not been previously employed on Finnish material, and the agglutinative nature of the Finnish language brought about its own challenges when strings of verbal text had to be divided into smaller sections for the purposes of analysis. In Section 4.2.3, I discuss how choice networks may be interpreted: what they were

originally proposed to reflect, how this changed in subsequent research and, finally, how I have chosen to interpret choice networks in my own research.

#### 4.2.1 Choice Networks: Choosing from Options

Choice Network Analysis (CNA) compares the translations of the same source text by multiple translators in order to construct models of the cognitive processing underlying translation. CNA is introduced both as an alternative and a complement to other research methods used for inferring these models, such as think-aloud protocols which, while being rich sources of information, are unable to access processing that is automatized (Campbell 2000, 30) or, one might add, otherwise beyond conscious attention. CNA builds on the premise that while we cannot directly observe cognitive processing, we can infer a model of this processing from its results, the translations (Campbell 2000, 33).

In CNA, a network-like flowchart, or a choice network, is created by comparing and classifying the different translation solutions produced by a group of translators for a particular source text segment. In this approach, the term *option* refers to the “set of possibilities available to a subject when faced with the translation of a specific item” (Hale and Campbell 2002, 18). The set of options offered by a specific source text segment is arrived at by laying out all of the different translation solutions produced for the segment by a particular group of translators. CNA assumes that these options – the possible translation solutions – were, in principle, available for the entire group, and that a choice was needed in order to select one of them (*ibid.*).

One should emphasize that the options that are arrived at when comparing the translations made by a particular group of translators will only reflect the options that the source text segment offered for the members of the particular group. The members of another group might interpret the segment differently. Further, as I write in Article 3, it is obviously possible that the translators considered other possible translation solutions before settling on a particular option. In my analysis, I take the view that the options, as the translation solutions that were actually employed by the translators, only represent the translation solutions that the particular translators considered to be the most appropriate for the given context, not all solutions that were available for the translators.

Campbell (2000, 38) considers CNA as theory-free in the sense that it may be used to test theories or hypotheses according to the aims of each research setting. In my multimodally-informed analysis, I aimed to determine if the multimodal source

text – the verbal text combined with the images – offered options that the verbal text on its own does not. In other words, in my analysis, I set out to assemble the range of options offered by a particular source text segment, and analyze whether these options corresponded to verbal or visual information, or a combination of these.

Campbell proposes that a choice network can be used to make general inferences about what went on in the translators' minds when translating a particular source text segment or, as Campbell puts it, "the processes that typically operate in particular type of subjects translating particular kind of texts under specific conditions" (Campbell 2000, 31). Campbell (2000, 32) suggests that these inferences can subsequently be used to construe more general principles about cognitive processing or to be used as hypotheses in the examination of other texts. Yet, one can ask how much we may really infer of this cognitive processing simply by examining the results of the processing (translations). As also pointed out by Nataša Pavlović, CNA only displays the various results of the translators' decision-making process but not the reasons behind these decisions (2007, 178–179).

As I write above, Campbell's original account of the method does not take into consideration that translators may have considered other possible translation solutions before settling with the final one. The method therefore falls short in elucidating what went on the translators' minds when translating, as only the result of this reasoning may be used in the analysis. My analysis aims to show that CNA is a clever tool for translation research: it encourages the close comparison of different translation solutions and hence invites us to contemplate on the motives behind the differences. Yet, in order for a study to make claims about the translators' motivation behind their solutions, it is highly recommendable to triangulate CNA with a complementary, introspective data collection method, such as think-aloud protocols, translation diaries or other types of verbal reporting by the translators' themselves.

Sharon O'Brien (2006, 116) points out that Campbell does not address the issue of what is the number of individual translations that a choice network can and should include. The number of research participants used in Campbell's own studies quoted above ranged between nine and eleven (Campbell 2000, 34; Hale and Campbell 2002, 20–26). From a practical perspective, I wish to point out that constructing networks of a significantly larger number of translations might be overwhelmingly time-consuming and the networks themselves might quickly turn out to be too complex to be analyzed in a productive manner. I would therefore suggest that the method is best suited for relatively small sets of translations, such as the data of the present study. As O'Brien asserts, if we only examine the translations produced by a limited

group of participants, our analyses can hardly be claimed to be conclusive (2006, 116). Making conclusive claims is not my intention, however. Instead, I wish to emphasize that, as my choice network analyses readily display, the same multimodal source text could be interpreted in various ways; even with such a limited number of research participants, variation of interpretation was evident.

#### 4.2.2 Creating a Choice Network

A choice network is network-like flowchart in which the translations of a particular source text segment are visually laid out in a particular order to facilitate analysis. Campbell (2000, 39) lays out three main principles for the creating a choice network. First, the network must be linguistically plausible and grammatically constituent. Second, the network must include every piece of data in the data – nothing can be left out simply because it differs from the rest. Third, the network must be optimally parsimonious and only include the minimum number of branches that will account for the data. I provide practical examples of these principles below.

The size of the analyzed source segment is determined by the aim of the research (Campbell 2000, 38): it can be an individual word (e.g. Hale and Campbell 2002, 21–22), a string of several words (e.g. Campbell 2000, 34) or even an entire phrase (e.g. Pavlović 2007, 164–166). For the purposes of my own analyses – in which the individual translation solutions are meticulously compared – the verbal source text was divided into relatively small segments; in fact, all of the networks introduced as examples in Article 3 represent either prepositional phrases or compound nouns. In some of these networks, each segment has been further divided into smaller items: either individual words or chains of a few words. In practice, I had to create this division on a case-by-case basis, taking into consideration the structure of both languages involved.

Figure 3 displays an example choice network from my analysis, representing the options offered by the source text segment “in the upper part of the equipment”, extracted from the phrase

“The weakly magnetic and non-magnetic particles are carried forward by the stream and eventually discharged from a tailings launder in the upper part of the equipment.”

The segment has been further divided into two items, “in the upper part” and “of the equipment”. My decision to divide the segment in this particular fashion comes down to the structure of the Finnish language: both of the prepositional phrases “in

the upper part” and “of the equipment” can be translated into Finnish using a single word only (e.g. *yläosassa* and *laitteen*). The Finnish grammar would therefore not have allowed for a division in another part of the segment (e.g. “in the upper part of” and “the equipment”). This corresponds to Campbell’s first principle for creating a choice network, namely that the network must be linguistically plausible and grammatically constituent.



**Figure 3.** An example choice network from the analysis of the data, representing the translations of the prepositional phrase “in the upper part of the equipment”.

The choice network in Figure 3 represents the options that the source text segment offered for the research participants. The options are displayed in Finnish in italics. Options that have been employed by more than one participant are marked with a corresponding number (e.g. *yläosassa* x2). When examining the options offered by the item “in the upper part”, we notice that they can be divided into two distinctively different groups: options that maintain the original meaning of the item (“upper part”) and options that display a radical change in meaning (“lower part”). Including the radically different options in the network corresponds to Campbell’s second principle for creating the networks, namely that the network includes every piece of data in the data and nothing has been left out simply because it differs from the rest.

The fact that the solutions *alaosassa olevan* and *alaosan* (“located in the lower part” and “of the lower part”) have been paired up to a single category corresponds to Campbell’s third principle, that the network must be optimally parsimonious and only include the minimum number of branches that will account for the data. The same kind decision-making can be seen in other parts of the network as well (e.g. joining *yläosassa*, *yläosan*, and *yläosassa sijaitseran*, or “in the upper part”, “of the upper

part” and “situated in the upper part”, into one category). In constructing the network, I concluded that the translation solutions were similar enough to represent a certain type of solution made by several participants. These were, by no means, straightforward decisions to make, but required a number of iterative steps.

In my multimodally-oriented analysis, I compare these solutions to both the verbal and the visual part of the source text. In this particular example, the source text segment corresponds to a part in the multimodal source text in which a particular piece of information, the location of a *tailings launder*, is expressed in the two modes in contradictory ways. According to the verbal source text, the tailings launder is located in the upper part of the equipment while, according to the visual source text (see Figure 2), the part is one of the bottommost parts of the device. The analysis hence shows that two of the options offered by the item actually correspond to the visual source text.

As quoted above, the method is said to produce networks that may then be used to make inferences, to test theories and so on, but I propose that the actual creation of the network constitutes the major part of the analysis. It would obviously also be possible to compare a set of different translation solutions simply by presenting them as a list. There is no predetermined structure for a choice network – the layout needs to be decided by the analyst. Creating a network requires a thorough comparison of the different solutions and the ways in which they differ from each other. I suggest that a choice network reflects the analyst’s interpretation of what went on in translation and the structure of the network is guided by the aims of the analysis. There might seldom be only one possible way of producing a choice network of a particular set of data. In the example source text segment presented above, another analyst might have been interested in the linguistic means by which location is conveyed in the translation via the grammatical cases of the Finnish language, and the starting point of the analysis might have been to combine, for instance, *yläosassa olevan* with *alaosassa olevan* (“located in the upper part”, “located in the lower part”) into one category.

This reflection is not meant to downplay the importance of the observations I make in my choice network analysis. I propose that the method yields a useful tool for the empirical investigation of translation and I found it convenient for the purposes of my research questions. I simply emphasize the slightly subjective nature of choice networks since it is not explicitly mentioned in the related literature and it may go unnoticed by the readers of these studies. The upside of the method is that, as long as the readers of the research reports acknowledge that the aims of research may have guided the structures of the networks, the readers should have enough

information to judge whether the analyst's interpretation of the data has been plausible, as the networks explicitly present the entire data from a particular segment.

#### 4.2.3 Interpreting a Choice Network

I now move on to consider what CNA has been proposed to reveal about the characteristics of the source text being examined. Campbell (1999) originally proposed that counting the number of the different translation solutions produced for a specific source text item could be used to assess how *difficult* the item was to translate. I have not adopted this interpretation in my own analysis and I will now propose an alternative way of interpreting a choice network. Campbell's original rationale was that if a source text item was translated in the same way (or very few ways) by all translators of the group, it could be assumed that translating the item required relatively little cognitive effort and the item was therefore easy to translate. If, on the other hand, the item was translated in a number of different ways, it could be assumed that the range of possible translation solutions was available for all of the translators. Consequently, making a selection among these options would have required more cognitive effort and the item was difficult to translate.

Yet, in a later article, Hale and Campbell (2002) refined this original idea by comparing difficulty with translation accuracy. It is indeed possible to translate an item identically but inaccurately by every translator in a group. The choice network of such an item would hence consist of a single translation solution; yet, it would not be plausible to claim that the item was easy to translate. Hale and Campbell (2002, 29) therefore concluded that while it may sometimes be possible to assess the difficulty of a source text item by counting the number of the translation solutions produced, there is no clear correlation between the two.

The question has also been reflected on by Sharon O'Brien (2006, 46) who asks if a large number of different translation solutions could sometimes be attributed to translator creativity and individuality instead of source text difficulty. In her own research, O'Brien set out to examine the correlation between difficulty and the number of different renditions produced. This was done by comparing choice networks with Translog data, assessing the required cognitive effort by measuring hesitations and pauses during a particular task (2006, 55). O'Brien found what she considered tentative empirical evidence to support Hale and Campbell's original view: the parts of the source text that corresponded to complex choice networks had also required more hesitations and pauses than other parts of the text, thus reflecting

difficulty. However, one could obviously ask if we can be certain that hesitating and pausing while translating a specific item is necessarily indicative of difficulty – being creative might require pausing as well.

To consider the question of *difficulty* versus *creativity* further, let us imagine two different scenarios: In the first, a group of translators translate an abstract poem. In the second, a group of translators translate a medical text describing the implantation of a cardiac pacemaker. There is one item in both source texts for which the entire group produces a different translation (hence, complex choice networks). Here we are dealing with profoundly different text types and translation assignments of profoundly different nature. Perhaps we should not claim that the complexity of both choice networks is indicative of either source text difficulty or translator creativity and not the other. In order to find the common denominator behind the mutual complexity of these two choice networks, I propose that we need to examine the question from a slightly higher level of abstraction.

I suggest that, ultimately, we can only argue that a complex choice network reflects that the particular item is open for more possibilities of interpretation; that the meaning evoked by the item was, in one way or another, *indeterminate* or ambiguous. For some items – and for some translators – this might indicate that the item was difficult to translate (for a convincing example of this, see e.g. Hale and Campbell 2002, 25–27), sometimes it could indicate that the item offered the translators a chance to be creative. On other instances, this could indicate something unrelated to either of these. In the same fashion, the fact that the group of translators produced few or even a single translation solution for a particular item reflects that the item offered fewer options for interpretation: the meaning evoked by the item was more determinate or unambiguous for the group of translators. This way of interpreting a choice network does not necessarily need to comment on whether the translation solutions are accurate by any standards.

To sum up, I suggest that we can use choice networks to assess how determinate or indeterminate the meaning of the analyzed item was for a group of translators. I also suggest that the complexity or the conciseness of a choice network is of interest in itself even if it is not used to contemplate on the difficulty of the item in question. In my multimodally-oriented analyses, I have set out to analyze if the complexity of the networks could be indicative of word–image interaction and the complex layers of meaning that may emerge in the intricate interaction of the modes.

## 5 ANALYSIS OF PREVIOUS THEORIES OF WORD–IMAGE RELATIONSHIPS

In Article 4, I analyze previous research that theorizes word–image interaction and compare this to the empirical observations made in Articles 2 and 3. In other words, I examine how words and images have been theorized to interact in previous research and compare this to the way in which words and images appeared to interact in my own empirical research. In this chapter, I introduce the theoretical material I examined in the article and the procedure of analysis I followed when comparing the material to my empirical research. The main aim of this analysis was to examine if the previous theories of multimodal meaning construction hold their ground when compared to empirical observations made about how the combination of words and images was interpreted by the participants of my study. As discussed in Chapters 1 and 2, the rationale behind this analysis is that the translations may be examined as reflecting the research participants’ interpretations of the multimodal text, and that they hence offer empirical insights as to how multimodal meaning construction unfolded for these particular readers.

I start this section in Section 5.1 by discussing how word–image interaction has typically been theorized in previous research. Sections 5.1.1 and 5.1.2 introduce the two examples of previous research I scrutinized in my analysis, namely studies by Emily E. Marsh and Marilyn Domas White (2003) and Radan Martinec and Andrew Salway (2005). In the final Section 5.2, I introduce the procedure of analysis I followed when comparing the previous theories with my empirical analyses.

### 5.1 Modelling Word–Image Interaction in Previous Research

In previous research, a widely adopted way of modelling word–image interaction has been to propose a set of *word–image relationships* that may hold between words and images when the two modes are presented together. Defining a word–image relationship refers to describing the way in which particular piece of verbal information (for instance, a sentence) is assumed to interact with a particular piece of visual information (an image). Examples of this could include saying that

particular words and images *complement* each other, or *specify* each other. Typically, creating a classification of word–image relationships has been done in one of two ways. The first is to create a *taxonomy* or a classification list based on empirical observations that are arrived at by reviewing large amounts of material that combines words and images, such as children’s picturebooks or newspapers. The second is to select a certain set of structural rules governing the composition of verbal language, and to model word–image interaction on this structure.

In my analysis, I included an example representing each approach. The first is an example word–image relationship taxonomies: Emily E. Marsh and Marilyn Domas White (2003) offer a master taxonomy that combines a total of 24 taxonomies created by other authors and outlines a total of 49 possible relationships that may hold between words and images. The other example represents modelling word–image interaction on the grammar of verbal language: Radan Martinec and Andrew Salway (2005) model word–image relationships on the grammatical concept of logico–semantic relations, originally introduced to describe linguistic structures by M.A.K. Halliday (1985).

These two examples were chosen because both could be argued to be more comprehensive than similar classifications. Martinec and Salway’s classification has been complimented for being more exhaustive than others in John A. Bateman’s comprehensive review on grammatical classifications (2014, 197). Given that Marsh and White’s master taxonomy includes the ideas presented in 24 other studies, it is likely to be the most exhaustive taxonomy ever proposed. Both of these classifications claim to be *complete* in that they represent all relationships that can exist between words and images, and both claim to cover all illustrated text types (Martinec and Salway 2005, 343; Marsh and White 2003, 647–653). My analysis aims to examine whether these claims of exhaustiveness really hold when the classifications are compared to empirical insights, in other words, whether the classifications really do represent all possible relationships that may hold between words and images. I now introduce Martinec and Salway’s classification and the rationale behind it, and then move on to discuss Marsh and White’s master taxonomy and the 24 taxonomies that constitute it.

### 5.1.1 Word–Image Interaction as Logico–Semantic Relations

Various scholars interested in word–image relationships have selected a certain aspect of the composition of verbal language and used it as a template to reflect on

how words and images interact. Examples of such an approach include research conducted by Terry Royce (2002), modelling word–image interaction on the concept of *sense relations*, introduced for verbal language by Halliday and Hasan (1987), Maite Taboada and Christopher Habel (2013), modelling word–image interaction on Mann and Thompson's (1988) *rhetorical structure theory*, and Yu Liu and Kay L. O'Halloran (2009), modelling word–image interaction on Halliday and Hasan's (1985) concept of *texture*. Martinec and Salway (2005) have based their work on Halliday's concept of *logico–semantic relations*, referring to the nature of relationships between clauses in a *clause complex*, or a combination of clauses.

According to Halliday, the relationships between clauses in a clause complex fall in two main categories, namely *projection* and *expansion*. In Halliday's linguistic account of logico–semantic relations, projection refers to experience that has already been said or thought, expressed in reported speech or reported thought (1994, 252–253). For instance, in the clause complex, “*Will I make it through this?* the man wondered,” the quoted thought “*Will I make it through this?*” is a projected clause. In Martinec and Salway's (2005, 352, 355) multimodal account of logico–semantic relations, a projecting relationship between verbal and visual information can be found in speech and thought bubbles in cartoons: the projected clause is enclosed in a bubble.

Halliday's category of expansion refers to the different ways in which clauses may add information to other clauses, and it comprises three subcategories: extension, elaboration, and enhancement. Halliday's (1994, 225) concept of *extension* refers to a clause that adds new information to another clause (consider, for instance the clause complex, “*The man felt dizzy, and was taken to the hospital.*”) Halliday's (1994, 226) concept of *elaboration* refers to a clause that exemplifies or specifies another clause (for instance, “*The man was taken to the hospital, which was a good idea.*”) Finally, Halliday's (1994, 232) concept of *enhancement* refers to a clause that provides circumstantial information about time, place, cause or reason, and so on, to another clause (for instance, “*The man felt dizzy, because he had not eaten in days.*”)

In Martinec and Salway's framework, Halliday's ideas of expansion have been employed to characterize word–image relationships: Extension refers to a relationship between words and images in which one mode adds something new to the other (2005, 353). In elaboration, one mode makes the other more specific (*idem.*, 352), and in enhancement, one mode qualifies the other circumstantially by providing information about time, place, cause, and so on (*idem.*, 353).

In the analysis presented in Article 4, I do not intend to question whether these kinds of relationships can exist between words and images. Instead, my analysis reflects on whether these are the *only* kinds of relationships that can exist between

the two modes. Martinec and Salway's classification uses the logico-semantic relations of verbal language as a template on which word-image interaction is modelled. It can therefore only include the kinds of word-image relationships that seem to correspond to the logico-semantic relations of verbal language. If there are word-image relationships that do *not* correspond to these linguistic structures, these relationships would inevitably be left out of Martinec and Salway's classification. The same restriction applies to all classifications that model word-image interaction on the composition of verbal language.

### 5.1.2 Word-Image Interaction as a Taxonomy of Relationships

My analysis examines March and White's (2003) master taxonomy from the same perspective as the previous classification: reflecting on whether it provides a complete description of the relationships that may exist between words and images. Because of space constraints, Article 4 discusses Marsh and White's word-image relationship taxonomy without introducing and comparing the 24 taxonomies on which Marsh and White's research is based, so I will here briefly present my review of these prior studies. My analysis included reviewing these studies in order to assess if the word-image relationships proposed in this literature are represented in the master taxonomy in a comprehensive and representative manner.

Marsh and White (2003) compare and combine 24 taxonomies of word-image relationships developed in various fields of research. Four of the studies focus on children's literature, mainly picturebooks (Bodmer 1992; Fang 1996; Nikolajeva and Scott 2000; Schwarcz 1982). Four of the studies analyze dictionary design (Hancher 1996; Ilson 1987; Landau 1989; Zgusta 1989). Six of the studies represent education research, which have focused on the instructional value of images and the ways in which images may assist reading comprehension (Brody 1980; Duchastel 1978; Levin 1981; Levin and Mayer 1993; Levin, Anglin and Carney 1987; Woodward 1993). Four studies represent journalism research. These have examined multimodality in print newspapers (David 1998; Kress and van Leeuwen 1998; Wanta 1988) as well as television news that combine still images with spoken verbal text (Walma van der Molen 2001). Finally, five studies fall under a category Marsh and White call library and information design. These examine images in multimedia (Berenstein 1997), indexing images in databases for purposes of information retrieval (Hidderly and Rafferty 1997; Stam 1984), document design (Schrivener 1997), and the effect of the layout of words and images in instructional contexts (Peeck 1993). I was unable to

retrieve a copy of Robert Ilson's paper "Illustrations in dictionaries," which is a conference paper presented in 1985 and published in 1987, so this study was left outside my own analysis.

According to Marsh and White (2003, 650), their procedure of analysis for combining the taxonomies started by identifying the relationships described in each study of the data and combining the different variants of the same concept, in other words, the same kind of relationship described by slightly differing terms in different studies. This stage left the authors with 49 concepts or possible relationships between words and images. These concepts were then grouped in three main classes based on the "degree of interaction" between the modes: little relationship between words and images, close relationship between words and images, and images extending the meaning of the text (*idem.*, 651).

In my review of Marsh and White's background literature, I observed two types of difficulties related to juxtaposing the 24 studies. First, some of the authors may not have intended to provide a conclusive list of possible relationships that may hold between the modes, and, perhaps for this reason, have not provided explanations as to what exactly is meant by each proposed relationship. For instance, George R. Bodmer's (1992, 72) "taxonomy" has been extracted from a sentence in which he simply states that images serve to "expand, explain, interpret, or decorate a written text" in a children's book. Employing taxonomies that have not been explained in detail by their original authors, such as Bodmer's, becomes problematic when the proposed concepts (*expanding*, *explaining*, *interpreting* and *decorating*) are to be combined with concepts proposed by other authors. Since Bodmer did not define what he means by, for instance, an *interpreting* relationship between words and images, one should be careful about stating that this concept covers the same kind of multimodal meaning construction as a similar-sounding concept proposed by someone else (an *interpreting* relationship between words and images was also proposed by Joel Levin [1981] and Prabu David [1998], and these concepts have been combined in March and White's work).

The second observation I made was that some of the taxonomies do not really address the relationships that may hold between words and images. For instance, Paula Berenstein's (1997) article discusses the information value of images in multimedia. Berenstein describes the functions of images, saying that they can be *instructional*, showing the viewer how to do something, *explanatory*, illustrating processes or structures, and so on. Marsh and White have transferred these functions to the study of word–image relationships without acknowledging that the author does not, in fact, discuss word–image interaction.

In addition to these questions related to Marsh and White's background literature, when examining the actual master taxonomy, I observed two additional types of difficulties related to juxtaposing the 24 studies. Some of the 49 concepts in the master taxonomy are so similar that they might be hard to distinguish from each other. For instance, the taxonomy differentiates between relationships that are *condensing* ("reducing to essential elements"), *concentrating* ("reducing [unimportant] aspects of text") and *compacting* ("summarizing") (Marsh and White 2003, 670), and the difference between these very similar concepts is not elaborated.

The other type of difficulty is related to whether the proposed relationship types are internally coherent. To give an example, one of the 49 relationships presented in the taxonomy is called a *developing* relationship, described as a circumstance in which words and images "elaborate", "specify" or "amplify each other's meanings" (Marsh and White 2003, 671). This relationship type has been created by combining ideas presented by seven different authors. However, one might reasonably argue that specification of meaning and amplification of meaning do not refer to the same kind of meaning construction.

My analysis concluded that there are factors that might challenge Marsh and White's selection and treatment of previous taxonomies, and there are factors that might complicate the practical applications of the master taxonomy: some of the proposed relationships seem ambiguous and some seem repetitious. Yet, Marsh and White's master taxonomy is the most comprehensive taxonomy of word–image relationships that has been proposed. My analysis set out to examine if it is, indeed, complete in the sense that it covers all ways in which words and images may co-construct meaning.

## 5.2 Procedure of Analysis for the Comparison of Theoretical and Empirical Material

The comparison of the theoretical accounts of word–image relationships and my own empirical analysis started with going through the translation examples I examined in my choice network analyses of Article 3, and reflecting on these examples in the light of what the analysis of the translation diaries in Article 2 revealed about each example. I then summarized how multimodal meaning construction appeared to have taken place for the participants in each example, and compared this meaning construction to the two reviewed classifications to see if either one of them included a concept that could be used to describe it.

To offer an example of this procedure, I here present my analysis of how the participants translated the term *launder*, which appears in the source text in reference to two different parts of the devices. In mining industry related texts, the term would typically refer to a long narrow channel used to convey liquids in a vertical direction. However, in the source text images, the shape of the launders has been modified so that they do not correspond to their verbal definition: the launders are depicted as slightly wider than they are tall (see, for instance, the tailings launder in Figure 2).

I started the analysis by reviewing what the choice network analysis revealed about the translation solutions produced for *launder*. For this particular term, the participants produced four conclusively different translation solutions. Seven of them produced solutions that, in one way or another, diverged from the verbal original. Four out of the seven translated the term as *säiliö* (“container”), which is slightly problematic since “container” does not refer to a channel through which liquid passes: containers are used to *contain* something. I then compared these solutions to my analysis of the translation diaries, which suggested that the participants appeared to aim to make the translation solution match the visual depiction of the part of the device rather than make it match its verbal description in the source text. I concluded that, in this part of the multimodal source text, the visual information seemed to *distort* the participants’ interpretation of the verbal source text item.

I then moved on to reflect if and how this instance of multimodal meaning construction – the interpretation of the four participants who employed this particular solution – could be modelled by employing the two reviewed classifications of word–image relationships. This included going through the classifications looking for a relationship that would describe this kind of word–image interaction. For this particular example, I concluded that the interaction does not adhere to any of the word–image relationships proposed in Marsh and White’s taxonomy, or in Martinec and Salway’s classification: nothing is being projected and nothing is being added. I then did the same analysis for other translation solutions produced for the same term. To sum up, this procedure of analysis aimed to examine if the previous classifications of word–image relationships really cover all possible ways in which words and images can co-construct meaning, or if other kinds of relationships may also be possible.

## PART II

# **6 RESULTS**

This chapter summarizes the main results of my study. The sections correspond to the four research questions – and articles – of the study: each section presents the most important insights gained in answer to a particular research question. Section 6.1 presents my discussion on the theoretical basis of the study. Section 6.2 introduces the observations made in my phenomenographic analysis of the translation diaries. Section 6.3 discusses the Choice Network Analysis of the translations, and reflects on how the analysis of the translation diaries informed the analysis of the translations. Finally, Section 6.4 presents my observations on how the empirical analysis of Articles 2 and 3 could be used to inform our understanding of multimodal meaning construction outside of Translation Studies.

## **6.1 Cognitively Modelling the Translation of Illustrated Technical Texts**

The first research question of the study, examined in Article 1, was to analyze how the translation of an illustrated technical text could be cognitively modeled based on what has been proposed about translation as cognitive activity as well as what has been proposed about the cognitive processes involved in illustrated text comprehension. In particular, the article focused on the processes involved in source text comprehension and the mental representation constructed of the source language message. For this purpose, the article introduced two models that aim to explain the mental representations we create when reading illustrated texts, namely the Cognitive Theory of Multimedia Learning (CTML), developed by Richard E. Mayer, and the Integrated Theory of Text and Picture Comprehension (ITPC), mainly developed by Wolfgang Schnotz.

The models share some common features; for instance, both assume that the human mind processes information in two separate channels, the verbal and the visual, and that readers of illustrated texts actively select, organize and integrate

verbal and visual information in order to comprehend the text (Schnotz 2005, 57; Mayer 2005, 31). However, the models offer competing interpretations as to how verbal and visual information are integrated in the human cognitive system. CTML proposes that the verbal and visual information are integrated into a single mental representation consisting of both verbal and visual information (Mayer 2005, 40). ITPC, in turn, proposes that the cognitive system creates multiple representations – some verbal, some visual – which interact with each other (Schnotz and Bannert 2003, 147; Schnotz and Kürschner 2008, 180).

Regardless of which one of the models is deemed more accurate, the implications they offer for Translation Studies are that when translating an illustrated technical text, the translator's interpretation of the verbal text – and, consequently, the translation solution – may be shaped by visual information. The mental representation a translator constructs from an item of an illustrated source text can be negotiated from the combination of verbal and visual information. The article concluded that whether or not verbal and visual information are merged together or contested and compared, each unavoidably affects how the other is interpreted. It follows that the translation solution, too, can be built on information negotiated from two different modes.

Previous research introduced in the article established that images can have a negative effect on verbal text comprehension if they provide information that is contradictory to the verbal. The article hence proposed that contradicting visual information may also complicate the translator's efforts to arrive at an adequate translation solution. Previous research discussed in the article also emphasized that a person's interpretation of an image is affected by individual preferences and prior knowledge, and that each person, therefore, interprets images in a unique way. For this reason, the article proposed that the way in which a particular image might affect translation will depend on the individual preferences of the each translator.

In sum, the article suggested that when translating illustrated technical texts, translation solutions can be built on information negotiated from two different modes. Yet, the article emphasized that as translators themselves decide, either consciously or unconsciously, to what extent they process images, it is highly unlikely that multimodal integration of meaning would happen at every instance of translating segments of illustrated technical texts. The empirical parts of the study, introduced in the following two sections, set out to test these arguments.

## 6.2 Conceptualizing Word–Image Interaction during Translation

The second research question of the study, examined in Article 2, sought to characterize the ways in which the participants conceptualized the interaction of verbal and visual information when translating an illustrated technical text. This was done by conducting a phenomenographic analysis of the translation diaries produced by the participants. The analysis identified two distinct main ways, or main *categories*, which describe the different ways of conceptualizing the interaction: either making a conscious effort to read, interpret and translate the combination of the modes as whole (category A), or separating the two modes of the source text for comparison and evaluation (category B). These categories do not describe individual participants: the students expressed more than one way of conceptualizing word–image interaction when discussing different parts of the multimodal source text.

The descriptions characterizing category A – perceiving the combination of words and images as an inseparable whole – indicated a conscious effort to map visual and verbal information onto each other. The quotes presented in this category illustrated how translation solutions were negotiated from the interaction of the two modes; the verbal source text evoked various different translation solutions, and the image defined what the “the correct” solution was. When this negotiation was not possible due to information missing from the image, the students seemed to be left with an impression of not having understood the source text at all. In these descriptions, the students did not question the *truth value* of either mode. Neither was described as faulty in any way, despite the fact that the modes provided contradictory information. Instead, the students described their own comprehension capacity as faulty. In the group meeting, several students discussed the instances of contradiction in the source text and mentioned that it had not occurred to them that there could be a mistake in the source text – they had simply thought they were “too stupid” to understand the process correctly.

This suggests that the students’ translation process was disturbed by visual and verbal information challenging or contradicting each other. It is also worth mentioning that the one participant who reported not having looked at the images during the actual translation stage did not discuss these parts of the verbal source text in the translation diary. This could be said to reflect two things: firstly, the fact that the rest of the students indicated these parts as difficult to translate was indeed due to lack of coherence between the modes (as opposed to the verbal source text simply being more difficult in these parts of the source text). Secondly, this implies that a verbal text presented with images can be interpreted and translated differently

than the same text without images. The source text without the images is, in a sense, a different source text.

In category B – evaluating words and images as competing sources of information – three different stances were identified: participants either considered verbal text as more relevant than the images, considered them both as equally relevant, or considered the images as more relevant than the verbal text. In other words, the combination of words and images was approached with strategies reflecting varying attitudes towards the importance and relevance of both modes. It may be reasonably assumed that these underlying attitudes affect the way in which the multimodal source text will be translated: for instance, valuing images over words could lead to translation solutions being based more on visual information than the verbal. All in all, the analysis of the translation diaries concluded that the students did not merely assign the images to a decorative or contextual role; instead, they conceptualized them as an important source of information.

### 6.3 Examining Word–Image Interaction in the Translations

The third research question of the study, discussed in Article 3, was to examine if the participants' translation solutions reflected verbal information only, or a negotiation of meaning between verbal and visual information. This was done by conducting a Choice Network Analysis on the translations: comparing the translations of the same source text by multiple participants in order to empirically derive the options that were available when translating each verbal item.

The analysis of the translations strongly implied that the visual source text had affected the students' solutions in various ways. Some of the translation solutions in the data reflected the way in which the corresponding information was presented in the visual source text. Examples of these included choosing between singular and plural forms of nouns in keeping with the image, and choosing terminology that corresponded more to the way in which particular objects were depicted visually than to the way in which they were verbally described. In the most extreme cases, conflicting visual information could result in verbal information to being disregarded altogether, meaning that the students replaced a verbal description of an event with the way in which the event was presented in the image.

The article proposed that source text images are capable of reattributing the meaning of an element in the verbal source text. Yet, making claims about how the interpretation of verbal information is affected by the visual means that we are

looking at the subject from a verbally informed perspective – the interpretation of the image is obviously also affected by verbal information. One might therefore argue that interpreting the combination of words and images may result in a novel interpretation of the whole: verbal and visual information may redetermine each other's meanings. Moreover, the results of the analysis emphasized that the combination of verbal and visual information can be interpreted in various different ways: the meanings the two modes produce together are neither predetermined nor clear-cut. Further, the analysis clearly showed that multimodal integration of meaning does not always take place even if various modes are presented simultaneously: the translator does not necessarily acknowledge both modes with equal attention at all times.

My analysis of the translation diaries greatly benefited the analysis of the translations. For instance, in my analysis of a part of the source text in which the two modes provided contradictory information (see example choice network in Figure 3 on page 68), CNA only revealed that the majority of the participant conveyed the information expressed by the verbal text (“upper part”). Based on this information alone, one might suggest that the majority of the participants did not consider visual information at all when translating this part. Yet, the analysis of the translation diaries demonstrated that the majority of the translation students actually identified the verbal item as mistaken, but nonetheless insisted on conveying this information to their translations. In other words, without having additional data about the participants' own reflections, I would not have known that both of the modes of the source text were attentively scrutinized and that the choice between conveying the verbal or the visual information was negotiated with conscious attention.

## 6.4 Translation Research as a Means to Model Word–Image Interaction

The fourth research question of the study, discussed in Article 4, was to examine how translation research can be used to advance our understanding of multimodal meaning construction outside the immediate context of Translation Studies. I approached this by critically reviewing two studies that present a classification of possible word–image relationships, and comparing these to the empirical insights gained in Articles 2 and 3. In other words, I compared the way in which previous research has theorized word–image interaction to the way in which word–image interaction was reflected in the empirical data of my study. The first reviewed

classification is an example of word–image relationship taxonomies (Marsh and White 2003), and the second is an example of accounting for word–image relationships by modelling them on the structure of the grammar of verbal language (Martinec and Salway 2005).

In the article, I made two central claims. The first was related to the completeness of the reviewed classifications: both are claimed to be exhaustive and to cover all illustrated text types (Martinec and Salway 2005, 343; Marsh and White 2003, 647–653). However, I argued that there are more possible word–image relationships than the reviewed classifications propose. For instance, the classifications do not account for contradiction between word and image (unless this is purposefully executed in order to produce irony, for instance). As my empirical analyses readily showed, even if words and images are not designed into a coherent whole, the reader may still interpret them in relation to each other; the modes may still interact.

I proposed that this shortcoming of the classifications is perhaps due to the fact that the corpora on which the classifications are based only represent a narrow sample of the myriad of word–image combinations that we encounter every day. The corpora used in both studies consisted mainly of encyclopedias, advertising material and educational material, in other words, illustrated texts in which words and images have been carefully designed to work together. If these types of texts contain instances of contradiction of information between the verbal and visual modes, it usually serves a thoroughly calculated function, such as conveying humor and irony, or simply calling for the reader’s full attention, as is often the case with advertising.

However, due to the sheer production speed and volume of multimodal texts, the words and images of an illustrated text may not always be meticulously orchestrated to work together in the best possible way. To cite the technical translator and the technical communicator quoted at the very beginning of Part I, an operating manual of a new model of a device could go into print with images of the previous model, and a technical text might include an image that has been added merely for esthetic purposes. I concluded that that conceptualizing word–image interaction by modelling it on ready-made classifications may hamper us from perceiving the complexity of this interaction.

The second central argument presented in the article was that we cannot predetermine word–image interaction in a particular word–image pair in a conclusive way. This goes against what some other authors have proposed. For instance, John A. Bateman (2014, 190) admits that analyzing word–image interaction by comparing it to word–image relationship classifications can lead to situations in which different analysts describe the relationship in different ways. Bateman goes on to argue that

“such variability is *not* due to multimodal artifacts exhibiting a rich array of meanings to be drawn out” (2014, 190 [emphasis in the original]), but instead are due to the classification system not being developed enough. I argued that the opposite holds. The empirical analyses of my study can be used to demonstrate that multimodal objects may indeed exhibit various meanings and hence be interpreted in varying ways. This was evident in my empirical analyses: the participants approached the illustrated text with varying attitudes towards the importance of each mode and translated the text in ways that reflected diverse ways of interpreting the interaction of the modes. I concluded that translation research offers one possible means to model the interaction of the modes in an empirical manner.

# 7 DISCUSSION

In this final chapter, I reflect on the implications and limitations of my study. I start the chapter by contemplating on the contributions of my research. I first discuss the contribution of my research to Translation Studies, and then contemplate on how the research contributes to the wider, interdisciplinary field of multimodality studies. I then move on to reflect on the limitations of the research and the generalizability of the research results in Section 7.2. Finally, in Section 7.3, I discuss possible future avenues of research.

## 7.1 Theoretical and Practical Contribution

From the perspective of Translation Studies, the overarching aim of this study has been to further promote a multimodally-driven perspective into translation and to demonstrate that when we are dealing with illustrated material, translation should not be thought of as merely verbal activity. For most of the students examined in this study, images were integrally involved in the translation assignment. Images played a significant role in source text comprehension and they defined the way in which elements of the verbal text were translated. Visual information was also employed in what could be called pre- and post-translation stages: the source text comprehension stage as well as the target text revision stage. All in all, the empirical analysis of the study suggested that the students often seemed to rely on images to define the contents of words. While this can also be a logical thing for a translator to do, I emphasize that translators should possess the critical ability to evaluate and assess the role of visual information in the multimodal context. This also has implications for the practice of translation training: translation curricula should train future translators to be able to examine images with equal detail as they examine the verbal dimension of the documents they translate.

Illustrations can have a significant effect on verbal text comprehension. The analysis of the translation diaries implied that the translation students, as readers of the multimodal text, made an explicit effort to map visual and verbal information onto each other. When this was impeded by deleting information from the image,

they seemed to be left with an impression of not having understood the verbal text at all. The analysis of the diaries also reflected that images, as a means of representation, can be given surprisingly high truth value – that they can be regarded as truthful representations of reality even when there is something clearly illogical in the image. The analysis suggested that this type of details in the image were often explained by doubting one's own sense rather than questioning the accuracy of the image. It was also striking that, due to the conflict between verbal and visual information, many of the students considered one part of the verbal source text to be erroneous (see page 84), but still insisted on not correcting it in their translations. These observations could be considered to have implications for translation didactics also on a more general level: whether source texts are multi- or monomodal, the students should be trained to critically analyze source texts and to trust their own interpretations.

The most notable contribution this study can offer for interdisciplinary research examining word–image interaction is to emphasize that the classifications of possible relationships that we often rely on in our analyses do not cover all of the possible ways in which words and images can co-construct meaning. As I have aimed to demonstrate in Article 4, these classifications tend to be based on corpora consisting of non-comprehensive sets of data which do not necessarily represent the reality of multimodal text production, since the corpora only tend include examples of illustrated texts in which the two modes have been carefully orchestrated into a harmonious entity. As my research has aimed to show, readers can display attempts at integrating verbal and visual information even if the messages conveyed by the two modes are completely contradictory. Further, my research has emphasized that the integration of verbal and visual information might not always take place: the reader does not necessarily acknowledge all of the modes with conscious attention at all times, even if the modes are simultaneously made available for inspection. As obvious as this may sound, it is seldom explicitly taken into account in research examining word–image interaction.

As to the overall research approach adopted in the study, I have aimed to demonstrate that instead of analyzing a multimodal text per se, we may also conduct research into multimodal meaning construction by examining products that, in one way or another, reflect the readers' interpretation of the interaction of the modes. The crucial benefit of such an approach is that it enables a comparison of how different readers construct meaning from the combination of the modes. The research approach adopted in this dissertation hence corresponds to the way in which, for instance, Holsanova (2012, 252) has urged multimodally-oriented research

to develop: it takes into account that different people interpret the combination of word and image in different ways, and aims to describe multimodal meaning construction by emphasizing that illustrated texts are open to various interpretations. For this reason, my study has the potential to make a contribution to a variety of research fields by participating in developing novel theoretical perspectives into multimodality.

## 7.2 Limitations of the Study

In this section, I reflect on the factors in my research design that influence the interpretation of my findings. I start by discussing the data collection and analysis methods and then move on to reflect on how the generalizability of the observations might be constrained by the student status of my research participants and the size of the participant group.

The translation diaries turned out to be a handy, yet slightly non-comprehensive set of data. The collection of the diaries was relatively effortless – especially in a university in which the translation students write translation diaries for all of their assignments and are trained and accustomed to creating them. As the diaries were already in a written form and were submitted electronically, no additional preparation was needed before analysis apart from anonymization. Not instructing the students to reflect on the images in their diaries was a justifiable solution, as it was important for the aims of the research that the students commented on the role of images exactly as they intuitively would. Yet, precisely for this reason, the students' reflection on the roles of images was not likely to be comprehensively thorough. In my research, a group meeting was conducted a week after the assignments were handed in, but as the aim of this meeting was mainly to provide the students with an opportunity to receive feedback on their work, the students were not systematically consulted for more comprehensive reflections regarding word–image interaction. In subsequent research designs, diary writing could be complemented with a thorough interview with the translator(s), preferably quite soon after completing the translation assignment, so the reasons behind their solutions would still be fresh in the memory.

Despite the slightly limited nature of the diary data, the phenomenographic analysis of the diaries produced some rather interesting insights about the way in which the students discussed their conceptions of word–image interaction. The method proved to be a convenient way of highlighting the variation in the ways of

conceptualizing the phenomenon. The proposed categories need to be understood as reflecting the researcher's own interpretation of the data, but I suggest that the example quotes from the data serve to justify the interpretations I have made.

The Choice Network Analysis, too, worked well as an analysis method, even though I did find it necessary to revise some of the underlying assumptions behind the method. First, I proposed that the *options*, the translation solutions a particular group of translators ended up using in their translations, do not necessarily reflect *all* of the possible translation solutions that were available to the group as proposed by Hale and Campbell (2002, 18). Instead, they may only be claimed to reflect the solutions that the translators found most appropriate for the given context. Second, I questioned Campbell's claim about the method being able to provide insights to the cognitive processing behind the translations. As became evident when comparing my choice network analyses to the analyses of the translation diaries, CNA only displays the final results of translators' decision making, but offers no insights into how they were arrived and why. Third, I questioned the way in which the complexity of a choice network may be interpreted. In addition to choice networks possibly being indicative of source text difficulty (as proposed by Campbell 1999), or translator creativity (as suggested by O'Brien 2006), I proposed that on a more general level, we might assume that complex choice networks might reflect that the particular item is open for more possibilities of interpretation; that the meaning evoked by the item was, in one way or another, *indeterminate* or ambiguous. My multimodally-oriented analyses suggested that the complexity of the networks may sometimes be indicative of word–image interaction and the complex layers of meaning that may emerge in the intricate interaction of the modes.

The research participants of my study were translation students in the final stages of their translation training. One may therefore reflect on whether the observations made in my research can be generalized to apply to professional translators as well. If the source text of this study was given to a translator who is experienced in translating texts of this particular substance matter, it is very likely that the knowledge of the operating process and relevant terminology would decrease the need to resort to visual information for reference. As established in Section 3.1, the more we know about the subject domain of an illustrated technical text, the less we resort to the images to support our verbal text comprehension. Naturally, the same applies for translators as well. It is therefore likely that a group of translators with subject domain specific background knowledge would have relied on the images less than the students did and that, consequently, their translation solutions would not have

reflected the visual dimension of the source text as much as the students' solutions did.

As to the contribution of my study to interdisciplinary research into multimodality, however, I propose that the generalizability of my research results is not affected by the student status of the research participants. Research into multimodal text comprehension can validly examine how subjects with limited background knowledge of a particular subject domain interpret word–image interaction in an illustrated text. In Educational Psychology, for instance, most studies examining word–image interaction in illustrated technical texts have examined university students from various major programs and with differing levels of subject domain-related background knowledge, including students with little or no previous knowledge of the topic (e.g. Mayer and Gallini 1990; Hegarty, Narayanan and Freitas 2002; Schnottz, Bannert and Seufert 2002; Schnottz and Bannert 2003; Schnottz and Kürschner 2007).

This study was conducted with a somewhat limited group of eight research participants. It is important to acknowledge that a larger group of participants might have revealed different aspects about the topic in focus: additional participants would have been likely to produce translation solutions not represented in the data as it is now. As for the phenomenographic analysis of the translation diaries, including a larger number of research participants might have revealed yet other possible ways of conceptualizing the role of images during translation. However, the outcome spaces produced in phenomenographic analyses are never claimed to be exhaustive; they merely aim to be complete in the sense that nothing has been left out of the collective experience of a particular group under focus (Marton and Booth 1997, 125).

As for the choice network analyses of the data, one could argue that the group size was quite appropriate: conducting a choice network analysis of a considerably larger set of translations might result unproductively complicated, as discussed in Section 4.2.1. As for my discussion on multimodal meaning construction outside the immediate scope of Translation Studies, I suggest that the limited group size, in a way, further supports my arguments: I have proposed that the translations display various ways of interpreting the same combination of words and images. Even with a limited number of research participants, this variation of interpretation was evident.

## 7.3 Implications for Further Research

Despite the fact the essential role of images in technical translation is often recognized (e.g. Risku and Pircher 2004; Tercedor et al. 2009, 143; Byrne 2012, 26, 54; Olohan 2016, 54), fairly little research has been conducted on the topic. The possibilities for future research are therefore abundant. One possible line of future research would be to analyze translations made by experienced professional translators and to compare these two participant groups to see if translation experience and subject domain-specific background knowledge affect the way in which translators employ visual information. The data of the present study consisted of translations and translation diaries. Possible methods for data collection in future research could include in-depth interviews with the translators, as suggested above, or think-aloud protocols. Eye tracking research might provide interesting data as well, as they might be able to show if and how the translators' visual attention alternates between the visual and the verbal source texts.

I have emphasized that the observations made in this study regard word–image interaction in *technical* texts. As established in Section 2.1, different types of images are likely to affect our illustrated text comprehension in different ways. A possible, intriguing line of future research would be to perform similar analyses on translations of other illustrated text types, and compare these analyses. I have elsewhere (Ketola 2018) examined translation students' diaries written about the translation of children's picturebooks, which is another example of a text type in which word–image interaction is essential in the overall comprehension of the multimodal message. The data used in the picturebook study is not directly comparable with the translation diaries examined in the present dissertation – the assignment was instructed slightly differently and the students had less translation experience than the research participants of this dissertation did. Yet, a tentative comparison even on a very general level would reveal quite a different approach to images as a source of information. The students appeared to regard picturebook illustrations as decorative elements with little or no truth value (*idem.*, 198). One might propose that examining images from such a different underlying stance might mean that the way in which images are involved in translation is different in picturebook translation. Yet, as said, making the comparison properly would require a more coherent research design.

An important suggestion for future research stems from the advances that research on human cognition has witnessed in recent years. My study built its main arguments on the Cognitive Theory of Multimedia Learning and the Integrated

Theory of Text and Picture Comprehension, which are both based on the view of human cognition as information processing. The same applies for the majority of the theoretical frameworks that have been adopted in studies on translation as cognitive activity (O'Brien 2010, 2), until very recently, one might add. However, it is quite safe to say that this view of human cognition is rapidly becoming outdated. In recent years, theoretical and empirical developments in various disciplines have contributed to the rise of the so-called situated cognition hypothesis (for a comprehensive review on research, see e.g. Roth and Jornet 2013). This hypothesis views cognition as embodied, distributed and fundamentally social. Further – and, perhaps, most importantly for the present discussion – the hypothesis suggests that cognition often works without representations. Translation research has also reacted to these advances. For instance, Risku and Windhager (2013) and Risku (2014) suggest that translation process research needs to refocus its scope to examine translation as embodied and situated activity.

As to the incisiveness of the discussion presented in my research, in Article 1 in particular, I may conclude that it is based on what was the most comprehensive research results available at the time of writing this article. Yet, in the years to come, our working hypotheses about translation as cognitive activity as well as our working hypotheses about how illustrated texts are interpreted are bound to change drastically. Advancing our understanding of these phenomena is a compelling avenue for future research. These advances will surely also open up a wealth of opportunities for inquiry regarding translation and multimodal meaning construction and offer a basis on which we can model a situated account of the translation of illustrated texts.

Word–image interaction has received great research interest in a variety of disciplines; yet, this research tends to be empirically unsubstantiated. In this dissertation, I have proposed that research into translation offers one possible way to empirically exemplify how the combination of words and images may be interpreted. Yet, research into multimodality faces an urgent need for further empirical work, preferably from methodological positions that acknowledge the variation in the ways people interpret the interaction of modes.



# REFERENCES

- Åkerlind, Gerlese S. 2005. "Variation and commonality in phenomenographic research methods." *Higher Education Research & Development* 24 (4): 321–334.
- Alsop, Graham, and Chris Tompsett. 2006. "Making sense of 'pure' phenomenography in information and communication technology in education." *Research in Learning Technology* 14 (3): 241–259.
- Angelone, Erik, Maureen Ehrensberger-Dow, and Gary Massey. 2016. "Cognitive Processes." In *Researching Translation and Interpreting*, edited by Claudia V. Angelelli, and Brian James Baer, 43–57. London and New York: Routledge.
- Ashworth, Peter, and Ursula Lucas. 2000. "Achieving Empathy and Engagement: a practical approach to the design, conduct and reporting of phenomenographic research." *Studies in Higher Education* 25 (3): 295–308.
- Baradaran Hezaveh, Mina, and Gholam-Reza Abbasian. 2015. "Integrated Problem and Decision Reporting Retrospection into the Translation Process." *Journal of Applied Linguistics and Language Research* 2 (4): 12–27.
- Barbour, Rosaline. 2007. *Doing Focus Groups*. London: Sage.
- Barnard, Alan, Heather McCosker, and Rod Gerber. 1999. "Phenomenography: A Qualitative Research Approach for Exploring Understanding in Health Care." *Qualitative Health Research* 9 (2): 212–226.
- Bateman, John A. 2014. *Text and image: A critical introduction to the visual/verbal divide*. London: Routledge.
- Bateman, John A., and Karl-Heinrich Schmidt. 2012. *Multimodal Film Analysis: How Films Mean*. London: Routledge.
- Berenstein, Paula. 1997. "Moving multimedia: the information value in images." *Searcher* 5 (8): 40–49.
- Bodmer, George R. 1992. "Approaching the illustrated text." In *Teaching Children's Literature: Issues, Pedagogy, Resources*, edited by Glen E. Sadler, 72–79. New York: Modern Language Association of America.
- Boeriis, Morten, and Jana Holsanova. 2012. "Tracking visual segmentation. Connecting semiotic and cognitive perspectives." *Visual Communication* 11 (3): 259–281.
- Brody, Philip J. 1980. "Research on pictures and instructional text: difficulties and directions." Paper presented at the annual meeting of the Association for Educational Communications and Technology (ERIC Document Reproduction Service ED 196 422).
- Byrne, Jody. 2006. *Technical translation. Usability strategies for translating technical documentation*. Dordrecht: Springer.
- Byrne, Jody. 2012. *Scientific and Technical Translation Explained*. Manchester: St. Jerome.
- Campbell, Stuart. 1999. "A Cognitive Approach to Source Text Difficulty in Translation." *Target* 11 (1): 33–63.

- Campbell, Stuart. 2000. "Choice Network Analysis in Translation Research." In *Intercultural Faultlines. Research Models in Translation Studies I. Textual and Cognitive Aspects*, edited by Maeve Olohan, 29–56. Manchester: St. Jerome.
- Chesterman, Andrew. 2007. "On the Idea of a Theory." *Across Languages and Cultures* 8 (1): 1–16.
- Chueasuai, Pasakara. 2013. "Translation shifts in multimodal text: a case of the Thai version of Cosmopolitan." *JoSTrans, The Journal of Specialised Translation* 20: 107–121.
- Cibangu, Sylvain K., and Mark Hepworth. 2016. "The uses of phenomenology and phenomenography: A critical review." *Library & Information Science Research* 38: 148–160.
- Conde, José T. 2008. "Proceso y resultado de la Evaluación de Traducciones." PhD dissertation, University of Granada, Spain.
- Connors, Sean P. 2013. "Weaving multimodal meaning in a graphic novel reading group." *Visual Communication* 12 (1): 27–53.
- Damaskinidis, George. 2015. "Mediating between verbal and visual semiotic elements in the translation of English multimodal texts into Greek." *Punctum* 1 (2): 22–38.
- David, Prabu. 1998. "News concreteness and visual-verbal association: do news pictures narrow the recall gap between concrete and abstract news?" *Human Communication Research* 25 (2): 180–201.
- Duchastel, Phillippe C. 1978. "Illustrating instructional texts." *Educational Technology* 18 (11): 36–39.
- Englund Dimitrova, Birgitta. 2010. "Translation Process." In *Handbook of Translation Studies*, Vol. I, edited by Yves Gambier, and Luc van Doorslaer, 406–411. Amsterdam and Philadelphia: John Benjamins Publishing Company.
- Espindola, Elaine. 2012. "Systemic functional linguistics and audiovisual translation studies: a conceptual basis for the study of the language of subtitles." *DELTA: Documentação de Estudos em Lingüística Teórica e Aplicada* 28: 495–513.
- Fang, Zhihui. 1996. "Illustrations, text, and the child reader: what are pictures in children's storybooks for?" *Reading Horizons* 37 (2): 130–142.
- Feng, Dezheng, and Elaine Espindola. 2013. "Integrating systemic functional and cognitive approaches to multimodal discourse analysis." *Ilha do Desterro* 64: 85–110.
- Flewitt, Rosie, Regine Hampel, Mirjam Hauck, and Lesley Lancaster. 2009. "What are multimodal data and transcription?" In *The Routledge Handbook of Multimodal Analysis*, edited by Carey Jewitt, 40–53. London and New York: Routledge.
- Forceville, Charles. 1999. "Educating the eye? Kress and Van Leeuwen's Reading Images: The Grammar of Visual Design (1996)." *Language and Literature* 8 (2): 163–178.
- Forceville, Charles. 2007. "Book review: Multimodal Transcription and Text Analysis: A Multimedia Toolkit and Coursebook." *Journal of Pragmatics* 39 (6): 1235–1238.
- Foulsham, Tom, and Geoffrey Underwood. 2008. "What can saliency models predict about eye movements? Spatial and sequential aspects of fixations during encoding and recognition." *Journal of Vision* 8 (2): 1–17.
- Francis, Hazel. 1996. "Advancing phenomenography – Questions of method." In *Reflections on Phenomenography. Toward a Methodology*, edited by Gloria Dall'Alba, and Biörn Hasselgren, 35–47. Göteborg: Kompendiet.
- Gall, Meredith D., Joyce P. Gall, and Walter R. Borg. 2007. *Educational Research. An Introduction*. Boston: Pearson.

- Garavini, Melissa. 2014. "La traduzione della letteratura per l'infanzia dal finlandese all'italiano: l'esempio degli albi illustrati di Mauri Kunnas." PhD dissertation, University of Turku, Finland.
- Gibbons, Alison. 2012. *Multimodality, Cognition and Experimental Literature*. New York: Routledge.
- Gile, Daniel. 2004. "Integrated Problem and Decision Reporting as a Translator Training Tool." *Journal of Specialised Translation* 2: 2–20.
- Göpferich, Susanne. 2010a. "Data Documentation and Data Accessibility in Translation Process Research." *The Translator* 16 (1): 93–124.
- Göpferich, Susanne. 2010b. "The Translation of Instructive Texts from a Cognitive Perspective: Novices and Professionals Compared." In *New Approaches in Translation Process Research*, edited by Susanne Göpferich, Fabio Alves, and Inger M. Mees, 5–55. Copenhagen: Samfundslitteratur.
- Göpferich, Susanne, and Riitta Jääskeläinen. 2009. "Process Research into the Development of Translation Competence: Where Are We, And Where Do We Need To Go?" *Across Languages and Cultures* 10 (2): 169–191.
- Hale, Sandra, and Stuart Campbell. 2002. "The Interaction Between Text Difficulty and Translation Accuracy." *Babel* 48 (1): 14–33.
- Hall, Stuart. 1980. "Encoding/Decoding." In *Culture, Media, Language: Working Papers in Cultural Studies 1972–79*, edited by Stuart Hall, Dooty Hobson, Andrew Lowe, and Paul Willis, 128–138. London: Hutchinson.
- Halliday, M.A.K. 1978. *Language as Social Semiotic: The Social Interpretation of Language and Meaning*. London: Edward Arnold.
- Halliday, M.A.K. 1985. *An Introduction to Functional Grammar*. London: Edward Arnold.
- Halliday, M.A.K., and Christian M.I.M. Matthiessen. (2004) 2013. *An Introduction to Functional Grammar*. 3rd ed. New York: Routledge.
- Halliday, M.A.K., and Ruqaiya Hasan. 1985. *Language, text and context*. Victoria: Deakin University.
- Halliday, M.A.K., and Ruqaiya Hasan. 1987. *Cohesion in English*. London: Longman.
- Hancher, Michael. 1996. "Illustrations." *Dictionaries: Journal of the Dictionary Society of North America* 17: 79–115.
- Hansen, Gyde. 2003. "Controlling the Process; Theoretical and Methodological Reflections on Research in Translation Processes." In *Triangulating Translation*, edited by Fabio Alves, 25–42. Amsterdam and Philadelphia: John Benjamins Publishing Company.
- Hansen, Gyde. 2006. "Retrospection methods in translator training and translation." *Journal of Specialised Translation* 5: 2–41.
- Hansen, Gyde. 2013. "The translation process as object of research." In *The Routledge Handbook of Translation Studies*, edited by Carmen Millán, and Francesca Bartrina, 88–101. London and New York: Routledge.
- Hasselgren, Biörn, and Dennis Beach. 1997. "Phenomenography – a 'good-for-nothing brother' of phenomenology? Outline of an analysis." *Higher Education Research & Development* 16 (2): 191–202.
- Hegarty, Mary, and Marcel A. Just. 1993. "Constructing Mental Models of Machines from Text and Diagrams." *Journal of Memory and Language* 32 (6): 717–742.
- Hegarty, Mary, N. Hari Narayanan, and Pam Freitas. 2002. "Understanding Machines from Multimedia and Hypermedia Presentations." In *The Psychology of Science Text Comprehension*, edited by José Otero, José A. Léon, and Arthur C. Graesser, 357–385. Mahwah, NJ: Erlbaum.

- Heidegger, Martin. 1962. *Being and time*. New York: Harper and Row.
- Henderson, John M., James R. Brockmole, Monica S. Castelhano, and Michael Mach. 2007. “Visual Saliency Does Not Account for Eye Movements During Visual Search.” In *Eye Movements: A Window on Mind and Brain*, edited by Roger P.G. van Gompel, Martin H. Fischer, Wayne S. Murray, and Robin L. Hill, 537–562. Amsterdam: Elsevier.
- Hidderly, Rob, and Pauline Rafferty. 1997. “Democratic indexing: an approach to the retrieval of fiction.” *Information Services and Use* 17 (3): 101–109.
- Hiippala, Tuomo. 2012. “The Localisation of Advertising Print Media as a Multimodal Process.” In *Multimodal Texts from Around the World: Cultural and Linguistic Insights*, edited by Wendy L. Bowcher, 97–122. London: Palgrave.
- Hiippala, Tuomo. 2013. “Modelling the structure of a multimodal artefact.” PhD dissertation, University of Helsinki, Finland.
- Hirvonen, Maija. 2012. “Contrasting Visual and Verbal Cueing of Space: Strategies and devices in the audio description of film.” *New Voices in Translation Studies* 8 (1): 21–43.
- Hirvonen, Maija, and Liisa Tiittula. 2010. “A method for analysing multimodal research material: audio description in focus.” *MikaEL, Electronic proceedings of the KäTu symposium on translation and interpreting studies* 4: 1–12.
- Holsanova, Jana. 2012. “New methods for studying visual communication and multimodal integration.” *Visual Communication* 11 (3): 251–257.
- Holsanova, Jana, Nils Holmberg, and Kenneth Holmqvist. 2009. “Reading information graphics: the role of spatial contiguity and dual attentional guidance.” *Applied Cognitive Psychology* 23: 1215–1226.
- Horton, William. 2005. “Graphics: The Not Quite Universal Language.” In *Usability and Internationalization of Information Technology*, edited by Nuray Aykin, 157–187. Mahwah: Erlbaum.
- Ilson, Robert. 1987. “Illustrations in dictionaries.” In *The Dictionary and the Language Learner: Papers from the EURALEX Seminar at the University of Leeds, 1–3 April 1985*, edited by Anthony P. Cowie, 193–212. Tübingen: Max Niemeyer Verlag.
- Jääskeläinen, Riitta. 2000. “Focus on Methodology in Think-aloud Studies on Translating.” In *Tapping and Mapping the Processes of Translation and Interpreting*, edited by Sonja Tirkkonen-Condit, and Riitta Jääskeläinen, 71–82. Amsterdam and Philadelphia: John Benjamins Publishing Company.
- Jääskeläinen, Riitta. 2010. “Are all professionals experts? Definitions of expertise and reinterpretation of research evidence in process studies.” In *Translation and cognition*, edited by Gregory M. Shreve, and Erik Angelone, 213–227. Amsterdam and Philadelphia: John Benjamins Publishing Company.
- Jensen, Kristian T.H. 2010. “Distribution of Attention Between Source Text and Target Text During Translation.” In *Cognitive Explorations of Translation: Eyes, Keys, Taps*, edited by Sharon O’Brien, 215–236. New York: Continuum International.
- Jewitt, Carey. 2009a. “An introduction to multimodality.” In *The Routledge Handbook of Multimodal Analysis*, edited by Carey Jewitt, 14–39. London: Routledge.
- Jewitt, Carey. 2009b. “Different approaches to multimodality.” In *The Routledge Handbook of Multimodal Analysis*, edited by Carey Jewitt, 28–39. London: Routledge.
- Jewitt, Carey. 2014. “Multimodal approaches.” In *Interactions, Images and Texts. A Reader in Multimodality*, edited by Sigrid Norris, and Carmen Daniela Maier, 127–136. Boston: De Gruyter.

- Jiménez Hurtado, Catalina, and Silvia Soler Gallego. 2013. "Multimodality, translation and accessibility: a corpus-based study of audio description." *Perspectives. Studies in Translatology* 21 (4): 577–594.
- Kaindl, Klaus. 2004. "Multimodality in the translation of humor in comics." In *Perspectives on Multimodality*, edited by Eija Ventola, Cassily Charles, and Martin Kaltenbacher, 173–192. Amsterdam and Philadelphia: John Benjamins Publishing Company
- Kaindl, Klaus. 2013. "Multimodality and translation." In *The Routledge Handbook of Translation Studies*, edited by Carmen Millán, and Francesca Bartrina, 257–268. London: Routledge.
- Kelly, Peter. 2003. "Validity and discursive phenomenography." Paper presented at the Annual Conference of the British Educational Research Association, Exeter, September 12–14.
- Ketola, Anne. 2016a. "Metafunktioanalyysi kuvitettujen tekstien käänösanalyysissä" [Metafunction Analysis in the Translation Analysis of Illustrated Texts]. In *Kääntäminen, tulkkaus ja multimodaalisus. Menetelmiä monimuotoisten viestien tutkimiseen*, edited by Eliisa Pitkäsalo, and Nina Isolahti, 91–117. Tampere: Tampere Studies in Language, Translation and Literature, Series B.
- Ketola, Anne. 2016b. "Translating picturebooks: Re-examining interlingual and intersemiotic translation." *Proceedings of DRS 2016*, 1179–1190.
- Ketola, Anne. 2018. "Encounters with an Illustrated Text. An Analysis of Translator Trainees' Diaries." In *Translating Picturebooks: Revoicing the Verbal, the Visual, and the Aural for a Child Audience*, edited by Riitta Oittinen, Anne Ketola, and Melissa Garavini, 187–201. New York: Routledge.
- Ketola, Anne, and Roberto Martínez Mateo. 2018. "The Story of Ferdinand: Comparing the Spanish and Finnish versions." In *Translating Picturebooks: Revoicing the Verbal, the Visual, and the Aural for a Child Audience*, edited by Riitta Oittinen, Anne Ketola, and Melissa Garavini, 110–119. New York: Routledge.
- Kiraly, Donald C. 1995. *Pathways to Translation: Pedagogy and Process*. Kent: Kent State University Press.
- Kokkola, Sari, and Anne Ketola. 2015. "Thinking outside the "Methods Box": New Avenues for Research in Multimodal Translation." *Vakki Publications* 4: 219–228.
- Kress, Gunther, and Theo van Leeuwen. (1996) 2006. *Reading Images. The Grammar of Visual Design*. London: Routledge.
- Kress, Gunther, and Theo van Leeuwen. 1998. "Front pages: the critical analysis of newspaper layout." *Approaches to Media Discourse*, edited by Allan Bell, and Peter Garrett, 186–282. Malden: Blackwell Publishers.
- Kress, Gunther, and Theo van Leeuwen. 2001. *Multimodal discourse: the modes and media of contemporary communication*. London: Arnold.
- Krings, Hans P. 2001. *Repairing texts: Empirical investigations of machine translation post-editing processes*. Kent: Kent State University Press.
- Kussmaul, Paul. 2000. "A Cognitive Framework for Looking at Creative Mental Processes." In *Intercultural Faultlines. Research Models in Translation Studies I. Textual and Cognitive Aspects*, edited by Maeve Olohan, 57–72. Manchester: St. Jerome.
- Landau, Sidney I. 1989. *Dictionaries: The Art and Craft of Lexicography*. Cambridge: Cambridge University Press.
- Langacker, Ronald W. 1991. *Foundations of Cognitive Grammar. Descriptive Application*. Stanford: Stanford University Press.

- Larsson, Jan, and Inger Holmström. 2007. "Phenomenographic or phenomenological analysis: does it matter? Examples from a study on anaesthesiologists' work." *International Journal of Qualitative Studies on Health and Well-being* 2: 55–64.
- Levin, Joel R. 1981. "On functions of pictures in prose." *Neuropsychological and Cognitive Processes in Reading*, edited by Francis J. Pirozzolo, and Merlin C. Wittrick, 203–228. New York: Academic Press.
- Levin, Joel R., Gary J. Anglin, and Russell N. Carney. 1987. "On empirically validating functions of pictures in prose." In *The Psychology of Illustration, Volume 1: Basic Research*, edited by Dale M. Willows, and Harvey A. Houghton, 51–114. New York: Springer-Verlag.
- Levin, Joel R., and Richard E. Mayer. 1993. "Understanding illustrations in text." In *Learning from Textbooks: Theory and Practice*, edited by Bruce K. Britton, Arthur Woodward, and Marilyn Binkley, 95–113. Hillsdale: Lawrence Erlbaum Associates.
- Lincoln, Yvonna S., and Egon G. Guba. 1985. *Naturalistic Enquiry*. London: Sage Publications.
- Liu, Fung-Ming Christy. 2011. "On Collaboration: Adaptive and Multimodal Translation in Bilingual Inflight Magazines." *Meta* 56 (1): 200–215.
- Liu, Yu, and Kay O'Halloran. 2009. "Intersemiotic Texture: analyzing cohesive devices between language and images." *Social Semiotics* 19 (4): 367–388.
- Lybeck, Leif. 1981. "Arkimedes I klassen. En ämnepedagogisk berättelse" [Archimedes in the classroom. A subject-didactic narrative]. PhD Dissertation, University of Gothenburg, Sweden.
- Macmillan Dictionary. 2017. "Interaction." Macmillan Dictionary. Accessed February 21 2017. Available at: <http://www.macmillandictionary.com/dictionary/british/interaction>
- Mann, William C., and Sandra A. Thompson. 1988. "Rhetorical Structure Theory: Toward a functional theory of text organization." *Text* 8 (3): 243–281.
- Marsh, Emily E., and Marilyn Domas White. 2003. "A taxonomy of relationships between images and text." *Journal of Documentation* 59 (6): 647–672.
- Martinec, Radan, and Andrew Salway. 2005. "A system for image–text relations in new (and old) media." *Visual Communication* 4 (3): 339–374.
- Martínez-Mateo, Roberto. 2014. "The story of Ferdinand: De Nueva York a Salamanca." *Ocnos* 12: 25–56.
- Marton, Ference. 1981. "Phenomenography – Describing conceptions of the world around us." *Instructional Science* 10: 177–200.
- Marton, Ference. 1988. "Phenomenography: A Research Approach to Investigating Different Understandings of Reality." In *Qualitative Research in Education: Focus and Methods*, edited by Robert R. Sherman, and Rodman B. Webb, 141–161. London: The Falmer Press.
- Marton, Ference. 1994. "Phenomenography." In *The International Encyclopedia of Education*, edited by Torsten Husén, and T. Neville Postlethwaite, 4424–4429. 2nd ed. Oxford: Pergamon.
- Marton, Ference, and Shirley Booth. 1997. *Learning and Awareness*. Mahwah: Erlbaum.
- Marton, Ference, and Wing Yan Pong. 2005. "On the unit of description in phenomenography." *Higher Education Research & Development* 24 (4): 335–348.
- Mayer, Richard E. 1993. "Illustrations that instruct." In *Advances in Instructional Psychology*, edited by Robert Glaser, 253–284. Hillsdale: Erlbaum.

- Mayer, Richard E. 2002. "Using Illustrations to Promote Constructivist Learning from Science Text." In *The Psychology of Science Text Comprehension*, edited by José Otero, José A. Léon, and Arthur C. Graesser, 333–356. Mahwah, NJ: Erlbaum.
- Mayer, Richard E. 2005. "Cognitive Theory of Multimedia Learning." In *Cambridge Handbook of Multimedia Learning*, edited by Richard E. Mayer, 31–48. New York: Cambridge University Press.
- Mayer, Richard E., and Joan K. Gallini. 1990. "When is an illustration worth ten thousand words?" *Journal of Educational Psychology* 82 (4): 715–726.
- Merleau-Ponty, Maurice. 1962. *Phenomenology of perception*. London: Routledge.
- Merriam-Webster. 2017. "Imago". Merriam-Webster Dictionary. Accessed August 8, 2017. Available at: <https://www.merriam-webster.com/dictionary/imago>
- Millán-Varela, Carmen. 2004. "Exploring Advertising in a Global Context. Food for Thought." *The Translator* 10 (2): 245–267.
- Mills, Mark, Andrew Hollingworth, Stefan Van der Stigchel, Lesa Hoffman, and Michael D. Dodd. 2011. "Examining the influence of task set on eye movements and fixations." *Journal of Vision* 11 (8): 1–15.
- Morgan, David L. 2011. "Focus Group Interviewing." In *Handbook of Interview Research*, edited by Jaber F. Gubrium, and James A. Holstein, 141–159. Thousand Oaks and London: Sage.
- Mubenga, Kajingulu Somwe. 2009. "Towards a Multimodal Pragmatic Analysis of Film Discourse in Audiovisual Translation." *Meta* 54 (3): 466–484.
- Mubenga, Kajingulu Somwe. 2010. "Investigating norms in interlingual subtitling: a systemic functional perspective." *Perspectives: Studies in Translatology* 18 (4): 251–274.
- Muñoz Martín, Ricardo. 2010. "On paradigms and cognitive translatology." In *Translation and Cognition*, edited by Gregory M. Shreve, and Erik Angelone, 169–187. Amsterdam and Philadelphia: John Benjamins Publishing Company.
- Nikolajeva, Maria, and Carole Scott. 2000. "The dynamics of picturebook communication." *Children's Literature in Education* 31 (4): 225–239.
- O'Brien, Sharon. 2006. "Machine-translatability and post-editing effort: an empirical study using translog and choice network analysis." PhD Dissertation, Dublin City University, Ireland.
- O'Brien, Sharon. 2010. "Introduction." In *Cognitive Explorations of Translation: Eyes, Keys, Taps*, edited by Sharon O'Brien, 1–14. New York: Continuum International.
- Oittinen, Riitta. 2003. "Where the Wild Things Are. Translating Picture Books." *Meta* 48 (1–2): 128–141.
- Oittinen, Riitta. 1990. "The Dialogic of Text and Illustration: A Translatological Point of View." *TEXTconTEXT. Translation. Didaktik. Praxis* 1: 40–53.
- Oittinen, Riitta. 2004. *Kuvakirja kääntäjän kädessä* [Picturebook in the hand of the translator]. Helsinki: Lasten Keskus.
- Oittinen, Riitta. 2008. "Audiences and influences: Multisensory Translations of Picturebooks." In *Whose Story? Translating the Verbal and the Visual in Literature for Young Readers*, edited by María González Davies, and Riitta Oittinen, 5–29. Newcastle-upon-Tyne: Cambridge Scholars Publishing.
- Ojaniemi, Markku. 2013. "Changes in the Conceptualization of Leadership during an African College Course." PhD dissertation. University of Tampere, Finland.
- Olohan, Maeve. 2016. *Scientific and Technical Translation*. London: Routledge.
- Ottenheimer, Harriet Joseph. 2013. *The Anthropology of Language: An Introduction to Linguistic Anthropology*. 3rd ed. Belmont: Wadsworth.

- Oxford English Dictionary. 2010. “Technical.” *Oxford English Dictionary Online*, edited by Angus Stevenson. 3rd ed. Oxford: Oxford University Press.
- Parkin, Alan J. 2014. *Essential Cognitive Psychology*. Hove: Psychology Press.
- Patton, Michael Quinn. 2002. *Qualitative Research & Evaluation Methods*. Thousand Oaks: Sage.
- Pavlović, Nataša. 2007. “Directionality in Collaborative Translation Processes. A Study of Novice Translators.” PhD Dissertation. Universitat Rovira i Virgili, Tarragona.
- Peeck, Joan. 1993. “Increasing picture effects in learning from illustrated text.” *Learning and Instruction* 3 (3): 227–238.
- Pereira, Nilce M. 2008. “Book Illustration as (Intersemiotic) Translation: Pictures Translating Words.” *Meta* 53 (1): 104–119.
- Pérez-Gonzalez, Luis. 2014. “Multimodality in Translation and Interpreting Studies: Theoretical and Methodological Perspectives.” In *A Companion to Translation Studies*, edited by Sandra Bermann, and Catherine Porter, 119–131. Chichester: Wiley-Blackwell Publishing.
- Prinsloo, Paul, Sharon Slade, and Fenella Galpin. 2012. “Learning analytics: challenges, paradoxes and opportunities for mega open distance learning institutions.” In *Proceedings of the 2nd International Conference on Learning Analytics and Knowledge*, 130–133.
- Richardson, John T.E. 1999. “The Concepts and Methods of Phenomenographic Research.” *Review of Education Research* 69 (1): 53–82.
- Risku, Hanna. 2014. “Translation Process Research as Interaction Research. From Mental to Socio-cognitive Processes.” In *Minding Translation. Con la traducción en la mente, MonTi Special Issue 1*, edited by Ricardo Muñoz Martín, 331–353.
- Risku, Hanna, and Florian Windhager. 2013. “Extended Translation: A Socio-cognitive Research Agenda.” *Target* 25 (1): 33–45.
- Risku, Hanna, and Richard Pircher. 2008. “Visual Aspects of Intercultural Technical Communication: A Cognitive Scientific and Semiotic Point of View.” *Meta* 53 (1): 154–166.
- Roisko, Hilkka. 2007. “Adult Learners’ Learning in a University Setting: A Phenomenographic Study.” PhD dissertation. University of Tampere, Finland.
- Roller, Margaret R., and Paul J. Lavrakas. 2015. *Applied Qualitative Research Design*. New York: Guilford Press.
- Rosenblatt, Louise M. 1978. *The Reader, the Text, the Poem: The Transactional Theory of the Literary Work*. Carbondale: Southern Illinois University Press.
- Roth Wolff-Michael, and Alfredo Jornet. 2013. “Situated cognition.” *WIREs Cognitive Science* 4: 463–478.
- Royce, Terry. 2002. “Multimodality in the TESOL classroom: exploring visual–verbal synergy.” *TESOL Quarterly* 36 (2): 191–205.
- Rydning, Antin Fougner, and Christian Lachaud. 2010. “The reformulation challenge in translation context reduces polysemy during comprehension, but multiplies creativity during production.” In *Translation and Cognition*, edited by Gregory M. Shreve, and Erik Angelone, 85–108. Amsterdam and Philadelphia: John Benjamins Publishing Company.
- Sager, Juan C. 1994. *Language engineering and translation: consequences of automation*. Amsterdam and Philadelphia: John Benjamins Publishing Company.
- Saldanha, Gabriela, and Sharon O’Brien. 2013. *Research Methodologies in Translation Studies*. London: Routledge.

- Säljö, Roger. 1997. "Talk as data and practice - a critical look at phenomenographic inquiry and the appeal to experience." *Higher Education Research & Development* 16 (2): 173–190.
- Sandberg, Jörgen. 1996. "Are phenomenographic results reliable?" In *Reflections on Phenomenography. Toward a Methodology?*, edited by Gloria Dall'Alba, and Biörn Hasselgren, 129–140. Göteborg: Kompendiet.
- Schnotz, Wolfgang. 2005. "An Integrated Model of Text and Picture Comprehension." In *Cambridge Handbook of Multimedia Learning*, edited by Richard E. Mayer, 49–69. New York: Cambridge University Press.
- Schnotz, Wolfgang, and Christian Kürschner. 2008. "External and internal representations in the acquisition and use of knowledge: visualization effects on mental model construction." *Instructional Science* 36 (3): 175–190.
- Schnotz, Wolfgang, and Maria Bannert. 2003. "Construction and interference in learning from multiple representation." *Learning and Instruction* 13 (2): 141–156.
- Schnotz, Wolfgang, Maria Bannert, and Tina Seufert. 2002. "Toward an Integrative View of Text and Picture Comprehension: Visualization Effects on the Construction of Mental Models." In *The Psychology of Science Text Comprehension*, edited by José Otero, José A. Léon, and Arthur C. Graesser, 385–416. Mahwah: Erlbaum.
- Schriver, Karen A. 1997. *Dynamics in Document Design*. New York: Wiley Computer Publishing.
- Schwartz, Joseph H. 1982. *Ways of the Illustrator: Visual Communication in Children's Literature*. Chicago: American Library Association.
- Séguinot, Candace. 2000. "Management Issues in the Translation Process." In *Tapping and Mapping the Processes of Translation and Interpreting: Outlooks on Empirical Research*, edited by Riitta Jääskeläinen, and Sonja Tirkkonen-Condit, 143–148. Amsterdam: John Benjamins Publishing Company.
- Stam, Deirdre Corcoran. 1984. "How art historians look for information." *Art Documentation* 3 (4): 117–119.
- Stöckl, Hartmut. 2004. "In between modes: language and image in printed media." In *Perspectives on Multimodality*, edited by Eija Ventola, Cassily Charles, and Martin Kaltenbacher, 9–30. Amsterdam and Philadelphia: John Benjamins Publishing Company.
- Sulkunen, Pekka. 1990. "Ryhmähaastatteluiden analyysi" [Analyzing group interviews]. In *Kvalitatiivisen aineiston analyysi ja tulkinta*, edited by Klaus Mäkelä, 264–285. Helsinki: Gaudeamus.
- Susam-Sarajevo, Şebnem. 2002. "A 'Multilingual' and 'International' Translation Studies?" In *Crosscultural Transgressions: Research Models in Translation Studies II: Historical and Ideological Issues*, edited by Theo Hermans, 193–207. Manchester: St Jerome.
- Taboada, Maite, and Christopher Habel. 2013. "Rhetorical relations in multimodal documents." *Discourse Studies* 15 (1): 65–89.
- Tanner, Beth, and Pete Larson. 1994. "Worth a thousand words: choosing and using illustrations for technical communication." *Technical Communication* 41 (1): 150–153.
- Tercedor, Maribel, Esperanza Alarcón-Navío, Juan A. Prieto-Velasco, and Clara I. López-Rodríguez. 2009. "Images as part of technical translation courses: implications and applications." *Jostrans, The Journal of Specialised Translation* 11: 143–168.
- Tercedor-Sánchez, María Isabel, and Francisco Abadía-Molina. 2005. "The Role of Images in the Translation of Technical and Scientific Texts." *Meta* 50: 4.

- Torresi, Ira. 2007. "Translating the Visual. The Importance of Visual Elements in the Translation of Advertising Across Cultures." In *Across Boundaries: International Perspectives on Translation Studies*, edited by Dorothy Kenny, and Kyongjoo Ryou, 38–55. Newcastle: Cambridge Scholars Publishing.
- Tuominen, Tiina, Maija Hirvonen, Anne Ketola, Eliisa Pitkäsalo, and Nina Isolahti. 2016. "Katsaus multimodaalisen käännöstieteeseen" [Observations about Multimodality in Translation]. In *Kääntäminen, tulkkaus ja multimodaalisuus. Menetelmiä monimuotoisten viestien tutkimiseen*, edited by Eliisa Pitkäsalo, and Nina Isolahti, 11–24. Tampere: Tampere Studies in Language, Translation and Literature, Series B.
- UTA käännöskomenttiohje. [Unpublished instructions for writing translation diaries for the translation students at the University of Tampere].
- Van Meerbergen, Sara. 2010. "Nederländska bilderböcker blir svenska. En multimodal översättningsanalys" [Dutch picturebooks become Swedish. A multimodal translation analysis]. PhD dissertation, University of Stockholm, Sweden.
- Walma van der Molen, Juliette H. 2001. "Assessing text-picture correspondence in television news: the development of a new coding scheme." *Journal of Broadcasting and Electronic Media* 45 (3): 483–498.
- Walsh, Eleanor. 2000. "Phenomenographic analysis of interview transcripts." In *Phenomenography*, edited by John A. Bowden, and Eleanor Walsh, 19–33. Melbourne: RMIT.
- Wanta, Wayne. 1988. "The effects of dominant photographs: an agenda-setting experiment." *Journalism Quarterly* 65 (1): 107–111.
- Webb, Graham. 1997. "Deconstructing deep and surface: Towards a critique of phenomenography." *Higher Education* 33: 195–212.
- Woodward, Arthur. 1993. "Do illustrations serve a purpose in US textbooks?" In *Learning from Textbooks: Theory and Practice*, edited by Bruce K. Britton, Arthur Woodward, and Marilyn Binkley, 115–134. Hillsdale: Lawrence Erlbaum Associates.
- Wright, Sue Ellen, and Leland Wright. 1993. "Editors' Remarks: Technical Translation and the American Translator." In *Scientific and Technical Translation*, edited by Sue Ellen Wright, and Leland Wright, 1–9. Amsterdam and Philadelphia: John Benjamins Publishing Company.
- Youngs, Suzette, and Frank Serafini. 2013. "Discussing Picturebooks Across Perceptual, Structural and Ideological Perspectives." *Journal of Language and Literacy Education* 9 (1): 189–200.
- Zahavi, Dan. 2003. *Husserl's Phenomenology*. Stanford: Stanford University Press.
- Zgusta, Ladislav. 1989. "Ornamental pictures in dictionaries." In *Lexicographers and Their Works*, edited by Gregory James, 215–223. Exeter: University of Exeter Press.

## APPENDICES

Appendix 1. The source text used in the translation assignment.

## Chapter 8

# ORE BENEFICIATION

The term *beneficiation* refers to a variety of processes in which desired materials are concentrated from collected ore. The ore is crushed and ground into fine particles, and the valueless constituents are then removed using a suitable beneficiation method, such as electrostatic or gravity separation, flotation or magnetic separation.

### 8.1 Magnetic beneficiation process and equipment

The magnetic separation method is used when separating minerals with different magnetic properties. The method comprises of applying a magnetic field to a stream of ore particles in order to separate magnetic particles from non-magnetic particles. Magnetic separation can be performed in either a dry or wet environment, though the latter is more common. The following paragraphs introduce two different types of wet magnetic separation equipment, concurrent drum separators and counter-current drum separators.

#### The operating principle of wet magnetic separators

A wet drum separator consists of a non-magnetic rotating drum and a series of three to six stationary magnets inside it. The drum is submerged in a tank which follows the contour of the drum. The ore particles are fed into the separator in the form of slurry. Water is introduced into the separator to provide a current which keeps the ore particles in suspension.

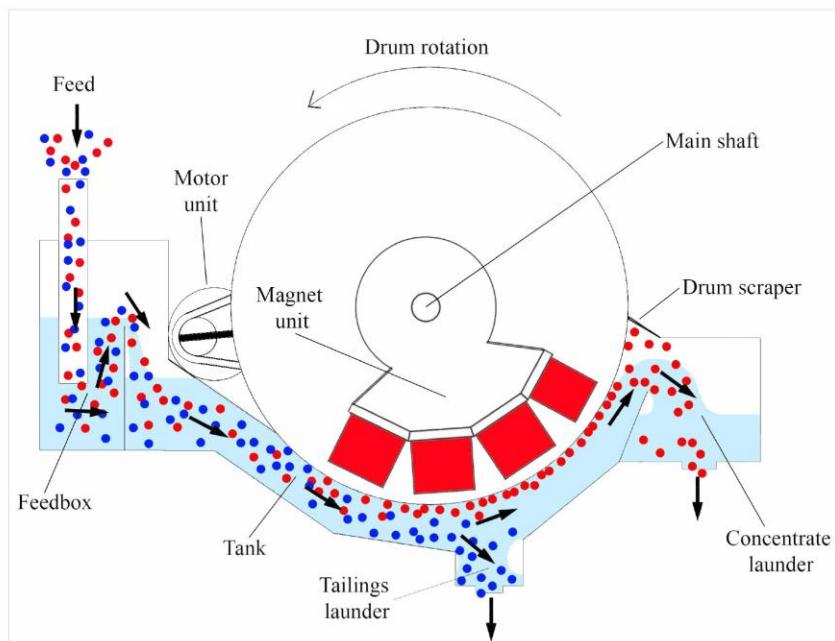
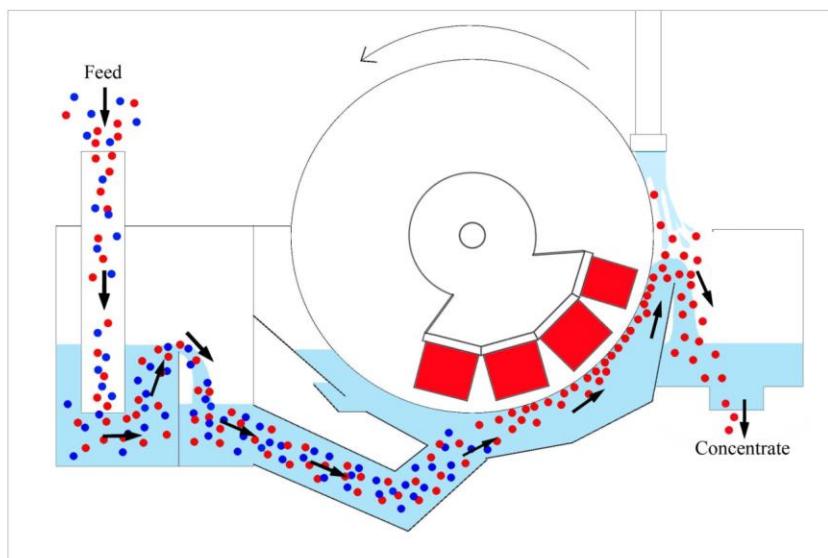


Figure 1. The parts and the operating principle of a concurrent drum separator

In a concurrent drum separator (Figure 1), the ore particles flow in the direction of the drum rotation. The slurry is first fed into the feedbox where it is diluted with water. It is then channeled under the drum where it enters the magnetic field generated by the magnets placed inside the drum. The rotational movement of the drum causes motion in the slurry, bringing solids into contact with the drum's surface. The magnetic particles are separated from the rest of the stream as they adhere to the drum surface in the area of the magnet. As the drum rotates past the magnet, the magnetic particles are conveyed out of the magnetic field and fall to a concentrate launder where they exit the separator. The weakly magnetic and non-magnetic particles are carried forward by the stream and eventually discharged from a tailings launder in the upper part of the equipment.

Figure 2 illustrates the operating principle of a counter-current drum separator. The term *counter current* is derived from the fact that the non-magnetic particles flow counter to the rotation of the drum when leaving the tank.



The slurry is introduced into the feedbox. It then proceeds through a narrow passageway near the bottom of the drum. When the slurry enters the magnetic field, the magnetic particles are attracted to the magnets and held against the surface of the drum. The rotation of the drum carries the particles out of the magnetic field, and the particles are washed off the surface of the drum with water sprays onto a concentrate launder. The non-magnetic particles flow in the opposite direction to the drum rotation and are discharged through an overflow into a tailings chute.

Appendix 2. The background information questionnaire given to the students (see English translation below).

Kyselylomake tutkimusta varten

MVK13E Tekniikan käänösviestinnän seminaari englanti-suomi  
Magneettirikastuskäännös

Opiskelijanumero: \_\_\_\_\_ Äidinkieli: \_\_\_\_\_

Arvioi pohjatietosi malmien rikastusprosessista asteikolla 1–5 (ympyröi kuvaavin vaihtoehto):

1. En tuntenut aihetta lainkaan.
2. Tunsin aiheen heikosti.
3. Tunsin aihetta jonkin verran.
4. Tunsin aiheen melko hyvin.
5. Tunsin aiheen erittäin hyvin.

Arvioi, kuinka hyvin uskot sisäistäneesi koneiden toimintaperiaatteet:

1. En ymmärrä toimintaperiaatetta lainkaan.
2. Ymmärrän toimintaperiaatteet osittain.
3. Ymmärrän toimintaperiaatteet jokseenkin.
4. Ymmärrän koneen toimintaperiaatteet melko hyvin.
5. Ymmärrän koneen toimintaperiaatteet hyvin.

Halutessasi voit kommentoida vastaustasi tähän:

---

---

---

---

---

Lomakkeen käänköpuolella on kopio käännettävästä tekstistä. Merkitse (esim. alleviivaamalla) tekstistä ne kohdat, jotka olivat mielestäsi muita hankalampia käänää. Halutessasi voit kommentoida vastaustasi alla oleville riveille tai käänköpuolelle käännettävän tekstin vireen.

---

---

---

---

---

---

English translation of the questionnaire.

Questionnaire for Research Project

MVK513E Technical Translation Seminar, English to Finnish

Magnetic Beneficiation Translation

Student number: \_\_\_\_\_ Native language:\_\_\_\_\_

Please assess your background knowledge in ore beneficiation prior to starting the assignment (circle the most appropriate option):

1. My background knowledge was very poor.
2. My background knowledge was poor.
3. My background knowledge was fair.
4. My background knowledge was good.
5. My background knowledge was very good.

Please assess how well you think you comprehended the operating principle of the devices:

1. I do not understand the operating principle at all.
2. I understand the operating principle poorly.
3. I understand the operating principle fairly well.
4. I understand the operating principle well.
5. I understand the operating principle very well.

You may comment on your reply below if you wish:

---

---

---

---

---

On the reverse side of the form, you'll find a copy of the source text. Please indicate (e.g. by underlining) which parts of the text you found particularly challenging to translate. If you wish, you may comment on the question either below or on the reverse side of the form next to the source text.

---

---

---

---

---

---

---

---

---

---

## PART III

# Towards a multimodally oriented theory of translation: A cognitive framework for the translation of illustrated technical texts

Anne Ketola\*

*School of Language, Translation and Literary Studies, University of Tampere, Finland*

\*Email: anne.ketola@uta.fi

This article introduces a cognitively grounded theoretical framework for the translation of illustrated technical texts, situating itself at the interface of translation studies and cognitive studies of illustrated text comprehension. The article proposes that translators of illustrated technical texts process both verbal and visual information and that, consequently, their translation solutions are built on information interpreted from the combination of two different modes. The article outlines a theoretical framework supporting these propositions by examining what has been proposed about translation as cognitive activity, and comparing this with two cognitive models of illustrated text comprehension, namely the Cognitive Theory of Multimedia Learning (e.g. Mayer 2002, 2005) and the Integrated Theory of Text and Picture Comprehension (e.g. Schnotz 2005). The article discusses the implications of the present discussion for future research on the translation of illustrated texts and emphasizes the importance of such research.

**Keywords:** Multimodality; technical translation; illustration; cognitive translation studies; cognitive models of multimedia learning

## Introduction

Multimodality, in essence, refers to the coexistence of more than one mode – written language, spoken language, images, and so on – within a given context (Gibbons 2012, 8). Multimodally oriented research posits that all modes present in a multimodal text contribute to meaning making. It emphasizes that while the analysis of verbal language is often an important part of the investigation of multimodal texts, verbal language is always embedded within (and interpreted in relation to) a wider multimodal context (Jewitt 2009a, 2). The consideration of multimodal issues is gaining importance in all research on communication including translation studies; after all, a large number of the texts being translated today are multimodal (Hirvonen and Tiittula 2010, 1). It is then only natural that, in recent years, translation scholars have urged the traditionally language-centered discipline to respond to these challenges (e.g. O'Sullivan 2013, 6; Kaindl 2013, 266).

Translation studies has already taken the first steps towards examining multimodal texts. Pioneering research contributions on the translation of illustrated texts include Kaindl's (2004) research into possible translation strategies for conveying humour in the translation of comics, and Oittinen's (e.g. 1990, 2000, 2008) work discussing picturebook translation strategies that take into account both verbally and visually presented information. Previous

research into the translation of illustrated technical texts, introduced in more detail below, has mainly focused on appropriate image selection and the need to adapt images during translation. More comprehensive research into the translation of illustrated technical texts is of great importance for two reasons: Firstly, technical translation is a significant area of translation, constituting a major share of the translation market (Byrne 2012, 6). Secondly, different types of images are a key feature of technical documents (Byrne 2012, 26, 54; Tercedor et al. 2009, 143). Before undertaking further research, however, preliminary theoretical framing of the subject from a cognitive perspective is in order.

The present article hence sets out to explore the cognitive implications of having the visual mode involved in the translation of technical texts. Multimodality as a research issue may naturally be approached from various perspectives. The perspective adopted by the present article could be described as cognitively grounded: it examines multimodal meaning making from the viewpoint of the cognitive processes involved in deriving meaning from various modes. The article builds on the premise that when reading an illustrated text, readers process both verbal and visual information, words and images, and form their interpretation of the multimodal text based on information provided by both of these modes. This notion has been established by investigating various types of illustrated texts, from children's picturebooks to scientific texts (and, logically, various types of readers from elementary school students to educated grown-ups), with methods ranging from reader interviews (Connors 2013; Youngs and Serafini 2013) and reading comprehension tests (see e.g. Mayer 2002 and Schnotz 2005 introduced below) to eye-tracking (Hegarty and Just 1993). Since translators of illustrated texts start their work as readers, this article proposes that the same argument applies for them as well. Consequently, it is proposed that their translation solutions are built on information interpreted from the combination of verbal and visual information. This, in turn, means that images may guide the way verbal text is translated.

The article provides a theoretical rationale for these claims by tentatively outlining the cognitive processes involved in illustrated text comprehension during translation, focusing on illustrated technical texts in particular. This is done by introducing two competing models of multimedia learning, namely the Cognitive Theory of Multimedia Learning (e.g. Mayer, 2002, 2005) and the Integrated Theory of Text and Picture Comprehension (e.g. Schnotz, 2005; Schnotz and Kürschner, 2008). The models have been constructed within the field of educational psychology and they describe the cognitive processes involved in illustrated text comprehension. They are based on the view of human cognition as information processing, and are therefore compatible with the majority of the theoretical frameworks adopted within cognitive studies of translation (O'Brien 2010, 2). The models are built on an extensive body of research, both theoretical and empirical, and have much to contribute to translation studies and our understanding of how the visual mode may be involved in the act of translation.

The purpose of the article is twofold. Firstly, the article seeks to contribute towards establishing a theoretical basis for multimodally oriented research in translation studies, and to extend the traditional boundaries of the discipline to include the examination of images as an object of inquiry in their own right. Secondly and most importantly, since the framework developed in the article articulates clear implications for further research, the article strives to push forward research questions hitherto ignored, encouraging others to further analyze, test and improve on these preliminary ideas. The article will begin by introducing previous

research efforts into the translation of illustrated technical texts and discussing what has been proposed about translation as cognitive activity. It will then move on to discuss how images are interpreted and how multimodal meaning making has been conceptualized in previous research into multimodality. It will then examine how readers construct an interpretation of an illustrated text by critically reviewing and comparing the cognitive models of multimedia learning. Lastly, the article will propose one possible interpretation for how the translator of an illustrated text may negotiate meaning from an illustrated technical text in the light of these two models and propose new pathways for empirical inquiry.

### **Technical and cognitive translation studies: Current state of research**

The definition of technical texts adopted by Franco Aixelá (2004, 32) includes “any text or text type in which there is a specific terminology belonging to a professional or academic field”. Byrne (2012, 26–28) describes the text type in a slightly more functionally oriented manner, including all task-oriented documents that seek to help a particular audience understand how to do something or how something works. When discussing *illustrated technical texts*, the present article focusses on informative, instructive texts that explain how something works and include visual information designed for the same function, examples of which would include user guides of physics textbooks. Technical texts such as technical data sheets, expert technical reports and certificates of conformity hence fall outside the research interest of the paper. The article hence builds on Byrne’s definition, keeping in mind that the field-specific terminology emphasized by Franco Aixelá is indeed a key feature of technical documents (Byrne 2012, 51).

Translation involves analyzing and considering a great number of factors. The need to analyze visual information in translation is acknowledged already in Christiane Nord’s model of text analysis (1991). The model advocates the analysis of images and other non-verbal elements together with other *intratextual factors* such as the subject matter, sentence structure and lexis, concerned with the contents of the text itself, as well as *extratextual factors* relating to the function of the source text, including the sender and the sender’s intention, the target reader, and so on (Nord 1991, 36–37). In fact, Nord (1991; 110) mentions technical texts such as manuals and operating instructions as examples of illustrated texts in which images play a particularly integral role.

Despite the importance of images in technical texts, fairly little has been said about the role of images in technical translation. In a 2007 overview of audiovisual translation, Remael and Neves briefly suggest that images in technical documents produce “new constraints, possibilities and translation problems that require increasingly creative solutions” (Remael and Neves 2007, 15), though they do not elaborate on what these could be in practice. The results of a teaching project (Tercedor et al. 2009, 144, 165) which aimed to familiarize technical translation students with image analysis confirmed that images indeed open up new possibilities for translators: the study indicated that accompanying the verbal source text with (task-appropriate) images led to more creative terminological translation solutions with less source language interference.

From a theoretical perspective, the translation of illustrated technical texts has previously been studied in detail only from the perspective of appropriate image selection. Tercedor-Sánchez and Abadía-Molina (2005) discuss the need to replace the images of the

source text during the translation process in order to ensure their proper transmission to a target audience, and provide general criteria for choosing images in technical and scientific texts (for discussion on similar issues within research into localization, see e.g. Hiippala 2012; Horton 2005). Prieto Velasco (2009, 2012), whose main interest lies in multimodal terminological databases, classifies technical and scientific texts and images based on their level of specification, explaining which types of images are most apt to represent technical and scientific concepts of varying levels of specification. In other words, these previous research efforts into illustrated technical texts have focused on the production of source texts (or database entries). The question of how these source texts are interpreted is yet to be assessed.

The issue of what happens inside the mind of the translator has been actively scrutinized over the past few decades and advanced by developments in relevant sub-disciplines of cognitive science as well as by new methods of data acquisition (for overviews, see e.g. Shreve & Angelone 2010, 2–9; Göpferich, Jakobsen & Mees 2009, 1–2). Cognitive translation process research sets out to elucidate the nature of the cognitive activities involved during the translation process, usually divided into source text comprehension, transfer between the two languages, and the production of the acquired information in the target language (Englund Dimitrova 2010, 406). The present article is not concerned with the translation process as a whole, but in the processes involved in source text comprehension and their result: in Jensen's (2010, 216–217) terms, the *mental representation* constructed of the source language message.

Advances in data acquisition methods within the field of cognitive translation studies include, for instance, key logging, used to examine the production processes in translation, and eye tracking, used to shed light on the source text reading and comprehension processes (Jensen 2010, 215–216); eye tracking data is generally assumed to be indicative of the cognitive processes involved (O'Brien 2006, 186). Eye tracking research has been used, for instance, to assess whether reading for translation differs from reading in monolingual contexts. Jakobsen and Jensen (2008) found that translators' reading time increases when they are told they will be asked to translate the text after reading it, as opposed to simply reading it for reading comprehension. This was interpreted as their reading being affected by some type of pre-translation. Similar observations were made by Jensen (2010), who concluded that, to some extent, source text comprehension and target text production are activated simultaneously when reading for translation. Further, Ruiz et al. (2008) found, albeit not with eye tracking methods, that reading for translation involves the activation of both lexical and syntactic matches in the target language. It may hence be concluded that reading for translation is a specific type of reading activity: whether consciously or unconsciously, it involves considering tentative translation solutions.

The act of reading a text results in an interpretation of its contents. One translator's interpretation of a source text is likely to diverge, even if only slightly, from that made by any other translator. As expressed by Muñoz Martin (2010, 175–176), “each act of understanding is unique and so are, consequently, translations”. In fact, Muñoz Martin emphasizes that one of the main tenets on which a cognitive theory of translation should be based is that it is not texts but individual interpretations that we translate. In other words, the object of translation is a unique interpretation, a mental representation, prompted by the source text segment and

actualized by reading comprehension processes. In what remains of the paper, I will provide theoretical support for the claim that when translating a multimodal text employing verbal and visual information, these mental representations are negotiated from the combination of the two modes.

### Modes making meaning

The perception of images in the cognitive apparatus may be divided into two phases: first, extracting information (shape, colour, size, depth) from the image and second, placing an interpretation on this information (see e.g. Eysenck and Keane 2005, Anderson 1995). Parkin (2014, 28) emphasizes that visual perception is *reconstructive*; our interpretation of an image does not develop in the mind as a photograph invariably develops in the emulsion of a film. Instead, the input from the retina of the eye provides a series of cues which are inferentially processed in order to create a somewhat unique internal representation of the image contents.

The process of “reading” images is far less straightforward than reading verbal text – a process that, by necessity, proceeds in a somewhat predetermined order. Some assumptions have traditionally been made about the order in which the viewer will scan the elements of an image. It is usually assumed that elements that are made salient or eye-catching by, for instance, colours, tonal contrast or relative size are inspected first. Further, the scan path is generally considered to be affected by the human tendency of moving one’s eyes in a particular reading direction (left to right in Western cultures), centering towards the center of the image (Foulsham and Underwood 2008), and focusing on human faces and their gaze directions (e.g. Lautenbacher 2012).

However, apart from the properties of the image, gaze is also directed by the viewer’s goals and expectations, emotions, prior knowledge and individual preferences (Boeriis and Holsanova 2012, 262). In fact, within studies of visual cognition, a growing body of research suggests that a person’s gaze is directed by cognitive information-gathering needs more than the inherent visual salience of the depicted scene (for reviews on research, see e.g. Mills et al. 2011, Henderson et al. 2007). In other words, viewing images is a task-oriented activity. These notions are of great importance for the current article, since they mean that different translators may attribute different meanings to the same image, and that the same translator may attribute different meanings to the same image depending on the purpose of viewing the image. Exactly the same claims were made about reading verbal texts in the previous section.

While it is challenging to comprehensively describe the meaning produced by individual modes in isolation, it is even more arduous to describe that created in the interaction of various modes. Despite the challenge, the issue has frequently been contemplated from various theoretical perspectives. The most prevalent and well-established of these are based on Halliday’s (1978) systemic functional linguistics, including social semiotic approaches (building on Kress & van Leeuwen 2001), which mainly focus on the metafunctions of modes, and the discourse analysis approach (e.g. O’Halloran 2008) which examines metafunctions in multimodal discourse at the micro-textual level. Even though the present article approaches multimodality from a cognitive perspective instead of concentrating on the metafunctions of modes, the conceptualization of word-image interaction is here adopted as a common conceptual ground for these differing approaches.

Hull and Nelson (2005, 225), building mainly on Kress and van Leeuwen, propose that the interaction of modes creates a whole new system of signification, transcending the combined contribution of individual modes. The idea is similar to that postulated by Lemke (2002, 303), who posits that the interaction of modes gives rise to genuinely new meanings and is therefore multiplicative in nature. O'Halloran (2008, 452) refers to this multiplicative interaction as intersemiosis, and asserts that it results in an expansion of meaning. Further, Lemke (2002, 303) has proposed that when presented together, the meanings of two modes may enhance and complement each other, offering specificity and precision beyond the capacity of either one alone. However, the meanings conveyed by individual modes in a multimodal text are not always perfectly aligned, but may also challenge or contradict each other. The relationships between modes may hence also create tensions between the aspects of meaning in a multimodal text (Jewitt 2009b, 26). The general idea proposed by these approaches therefore seems to be that when reading illustrated texts, the reader may infer meanings that are not to be found in the verbal text or the images alone; meanings that exist only in the combination of those particular words and those particular images – and for that particular reader and reading purpose, one might add. Depending on whether the messages conveyed by these two modes are aligned or not, the modes may either specify or challenge each other.

### **Reading illustrated technical texts**

As early as 1993, Hegarty and Just conducted a study in which they monitored the eye-fixations of test subjects reading an illustrated technical text. They found that the readers frequently interrupted their reading of the verbal text to inspect the image (an average of 6 times per page of 140 words). Instead of looking at the image at random, the readers fixated on the parts of the image that depicted the objects they had just read about in the verbal text (Hegarty and Just 1993, 730–731). While the findings are of utmost interest for research efforts such as the present one, the study offers no theoretical explanation as to what exactly occurs in the mind of the reader when visual and verbal information are integrated in this fashion.

Studies in educational psychology have since created cognitive models of multimedia learning that discern the process of illustrated text comprehension. The article will now introduce and compare the two most influential of these, namely the Cognitive Theory of Multimedia Learning (e.g. Mayer 2002, 2005) and the Integrated Theory of Text and Picture Comprehension (Schnottz, Bannert and Seufert 2002; Schnottz and Bannert 2003; Schnottz and Lowe 2003; Schnottz 2005; Schnottz and Kürschner 2008). Moreover, it will consider how the process of forming a mental representation of an illustrated source text would unfold in the light of these two models, and examine where the models stand in regard to empirical research findings. However, the section will begin by discussing issues of definitions. The applicability of the models of multimedia learning for the present article depends on their definitions of *multimedia* as well as their definition of *learning*; after all, one might fairly question if theories of multimedia learning are applicable to a framework of multimodal translation.

### ***Defining multimedia learning***

As stressed by, for instance, Kress and van Leeuwen (2001, 66–67), *multimediality*, when referring purely to the analysis of the medium, and *multimodality*, when referring purely to the analysis of the semiotic resources used to make meaning, are related yet independent concepts. However, different researchers and disciplines assign different definitions to the terms (see Lauer 2009 for an extensive review). For the purposes of their research, Mayer (2005, 32) and Schnotz (2005, 50) both define a multimedia message as a communication consisting of words (printed or spoken) and images (static or dynamic) delivered in any medium. While the definition is narrower than, for instance, Kress and van Leeuwen's (2006, 177) definition of a multimodal message – according to which a multimodal message consists of any two or more modes or “semiotic codes”, be they images, spoken or written language, music, gestures and so on – it unquestionably includes the subject of the present article, namely illustrated texts.

It is generally held that translators do not need to *learn* the content of the texts they translate, in the sense of being able to retrieve and apply the information after completing the translation task (e.g. Sager 1994, 199). Yet Mayer's definition of learning describes the process as applying “cognitive strategies to incoming information in order to make sense of it” (2005, 36). The outcome of this sense-making process is either one or several mental representations (see discussion below) stored in long-term memory. Neither of the models specifies how long these representations should be retrievable from long-term memory in order for the process to count as learning. The design of the research experiments used to test these theories reflects a view that the long-term retrievability of the acquired information is not of crucial importance: whether or not the test subjects learned the information presented to them is measured by tests of transfer (i.e. being able to use the information to solve new problems) immediately after reading it (Mayer and Gallini 1990, 719; Schnotz and Bannert 2003, 151; Schnotz and Kürschner 2008, 183; Schnotz, Bannert and Seufert 2002, 405). One could say that the models regard learning as apprehending the essential contents of something, which undoubtedly is a necessary prerequisite for successful translation activity as well.

Further, the authors behind the multimedia models do not explicitly specify which types of images their models account for. Yet, the images used in their research – as well as the research they quote – are technical in nature, describing, for instance, how lightning storms develop (Mayer 2002), how geographic time differences are determined (Schnotz and Kürschner 2008), or explaining the operating principles of pulley systems (Hegarty and Just 1993), pumps and car brakes (Mayer 2002). In other words, the models are built on analyses of precisely the type of illustrated technical texts that the article focusses on; therefore the models form a plausible basis for the present theory development.

### ***Models of illustrated text comprehension***

The Cognitive Theory of Multimedia Learning (hereafter CTML) and the Integrated Theory of Text and Picture Comprehension (hereafter ITPC) share some similar assumptions about the processes that unfold during the reading comprehension of illustrated texts. Both assume that the human mind processes information in two separate channels, the verbal and the visual, and that readers actively select, organize and integrate verbal as well as visual information in order to comprehend an illustrated text (Schnotz 2005, 57; Mayer 2005, 31).

However, the models differ in some of their basic hypotheses, the most notable being the phase of integrating verbal and visual information.

Mayer's CTML postulates that the reader proceeds through an illustrated text in small segments, selecting relevant words from the verbal text and selecting relevant parts of the image. This proposition seems quite plausible if we compare it to the observations made by Hegarty and Just introduced above. Their research subjects read a segment of verbal text and then inspected the part of the image that depicted the part of the process they had just read about. CTML suggests that the relevant verbal and visual information is then organized into a verbal model in the verbal channel and a visual model in the visual channel, respectively (Mayer 2005, 38–40). CTML proposes that the two separate models are then integrated by mapping corresponding elements and relations from one model onto the other, as well as connecting them with prior knowledge. In other words, the reader combines the information extracted from the two modes by making referential links between the two, resulting in a single, integrated representation (Mayer 2005, 40). Mayer views integrated representations as effective forms of information management and retrieval. One of the principal conclusions of Mayer's research is his “Multimedia Principle”, which asserts that illustrated texts may be more readily comprehended than texts with verbal information only (Mayer 2002, 352–353; 2005, 31–32).

If we were to use Mayer's model of illustrated text comprehension to elucidate the translation of an illustrated technical text, how would we perceive the process of source text comprehension? First of all, the translator would read small segments of the illustrated source text, selecting parts of verbal and visual information. The translator would then form a coherent representation of the selected verbal information and a coherent representation of the selected visual information, and then integrate these with the help of prior knowledge. The input for the particular translation task would hence be a single mental representation consisting of both verbal and visual information and the task of the translator would be to formulate a target language expression describing this representation. Based on the discussion on reading for translation introduced above, one could assume that the source text comprehension process also includes instances of pre-translation; the translator considering possible translation solutions for these mental representations whilst still reading the source text.

ITPC proposes an alternative interpretation for the process of reading comprehension. Most importantly, it questions the parallelism of processing words and images as described in Mayer's model, stressing that verbal text and images are based on different sign systems and employ different principles of representation: Verbal texts are descriptive representations consisting of symbols and are therefore associated with the content they represent by means of convention only. Images, on the other hand, are depictive representations consisting of iconic signs, associated with the content they represent through common structural characteristics (Schnotz and Kürschner 2008, 177). ITPC posits that when processing verbal text and images, the reader constructs internal mental representations that are also either descriptive or depictive. As these two forms of representation are fundamentally different, ITPC maintains that they cannot be mapped onto each other in order to create a single integrated model (Schnotz and Bannert 2003, 143; Schnotz and Kürschner 2008, 177).

Instead, ITPC proposes that the reader constructs multiple complementary representations both in verbal text comprehension and in image comprehension. Descriptive representations are referred to as *propositional presentations* and depictive as *mental models*. When reading verbal text, the reader constructs a descriptive propositional representation of the selected verbal text contents. This triggers the construction of a coherent depictive mental model, representing the typical visual features of what is described in the verbal text. When viewing an image, the reader constructs a depictive mental model of the selected visual information, as well as a complementing propositional representation describing the acquired information by verbal means (Schnotz and Bannert 2003, 145–146). These mental models and propositional representations are not integrated into a single representation, but interact with each other in a continuous process of what ITPC terms “mental model construction and inspection”. Each of them may be elaborated – or contested – by playing off new information from the others (Schnotz and Bannert 2003, 147; Schnotz and Kürschner 2008, 180).

According to ITPC, the translator of an illustrated technical text would again start by selecting segments of relevant visual and verbal elements. However, the number of representations formed during source text comprehension would differ from the previous model. When reading verbal text, the translator would construct both a propositional representation of the selected verbal text contents, as well as a coherent mental model depicting its typical features in visual form. In the same fashion, when retrieving visual information, the translator would build both a depictive mental model and a descriptive propositional representation verbally describing the typical features of the acquired information. The input for the translation task would consist of all of these representations which may both complement and contest each other. Again, the translator would not necessarily read the entire source text before considering translation solutions, but might consider translation solutions for one or even several of these representations as soon as they are constructed.

### ***Evaluation of the models***

A body of research in empirical psychology conducted in the late 1980s and early 1990s (e.g. Hegarty and Just 1989, 1993; Mayer 1989; Glenberg and Langston 1992) concludes that a text is easier to comprehend when it is illustrated, thus supporting Mayer’s Multimedia Principle introduced above. In these experiments, subjects reading an illustrated text comprehended and recalled the subject matter better than the control group reading verbal text alone. Such research findings have been interpreted as confirming that readers integrate the information in verbal text and images to construct an effective, integrated model of their common referent, rather than constructing separate representations of the two (Hegarty and Just 1993, 718), thus confirming Mayer’s idea.

However, all of these research settings have employed images whose content is aligned with that of the presented verbal text: the images are particularly designed to support verbal text comprehension in the best possible way. While the research findings are interesting and undoubtedly helpful for purposes of textbook design, for instance, it should be emphasized that these studies observe illustrated text comprehension from one limited perspective only. As repeatedly pointed out by Schnotz and his colleagues (Schnotz and Lowe 2003, 119; Schnotz and Bannert 2003, 153; Schnotz and Kürschner 2008, 181), CTML

ignores the fact that a subject may be visualized in more than one way. Schnotz and Bannert (2003) as well as Schnotz and Kürschner (2008) propose that not all forms of illustrating a text result in its advanced comprehension. Their research findings confirm this view. Images promote text comprehension when the image is designed to support the comprehension of the verbal text in the best possible way, but they may also have a negative effect on verbal text comprehension if the subject matter is visualized less appropriately for the given verbal context (Schnotz and Bannert 2003, 153–154; Schnotz and Kürschner 2008, 187–188). It should be emphasized that this observation further supports the premise on which the present article is based, namely that the interpretation constructed from an illustrated text is based on information provided by both modes. If this was not the case, there would be no reason why the appropriateness or inappropriateness of the image for the given verbal context should affect the way in which the verbal information is interpreted.

Therefore, while CTML is perhaps slightly more straightforward than ITPC, one should question its applicability in situations where the messages conveyed by the two modes are not perfectly aligned. In particular, one may question if a reader is able to form a single, consistent interpretation of verbal and visual information in cases where the two are clearly contradictory. Even though ITPC is better suited to account for these types of texts, this model does have its own weaknesses. For instance, Reimann (2003, 250) calls for clarifications particularly in the model construction and inspection phase of the model, which, for the purposes of the current article, is of most interest. Reimann also questions the empirical testability of the somewhat complex model, asserting that novel research designs would be needed in order to empirically support it, even though refraining from suggesting what these research designs could be like. Further, in recent years, there has also been general criticism of the information-processing view of the human mind on which these models are based (see Sorden 2013 for a review). For instance, the emotions involved in illustrated text comprehension are starting to generate research interest; yet the role of affect in the process is yet to be firmly established (Sorden 2013, 21). In conclusion, it is out of the scope and ability of any single article to decisively determine which one of the models most accurately describes the cognitive processes involved. Yet, regardless of which of the models is deemed more accurate, the implications of the above discussion for translation studies remain essentially the same: when translating an illustrated technical text, the translator's interpretation of the verbal text – and, consequently, the translation solution – may be shaped by visual information.

### **Conclusions and implications for future research**

The purpose of the present article was to discuss the cognitive implications of having images involved in the translation of illustrated technical texts. We might propose that the translator's reading comprehension process starts with constructing separate representations of the messages conveyed by the two modes. If the messages are aligned, building solid connections between the representations, each supporting the comprehension of the other, should be relatively easy. Therefore, the result could perhaps resemble a single, fused mental representation (as in CTML). If, however, the messages conveyed by the two modes are not aligned, building meaningful connections between the representations may not be possible. The comprehension process might then involve negotiating, debating, perhaps bargaining

between the information conveyed by the two modes (as in ITPC). The conclusion that emerges is that whether or not the verbal and the visual information conveyed by the source text are merged together or contested and compared, each unavoidably affects how the other is interpreted. It follows that the translation solution, too, is built on information negotiated from two different modes.

At a more particular level, the discussion presented in the article provides the following insights to inform future research:

*A verbal text coupled with images might be translated differently than the same text without images; and a verbal text coupled with one set of images might be translated differently than the same text with another set of images.*

Having concluded that the combination of particular words and images may give rise to meanings not to be found in either one alone, it may be postulated that adding illustrations to a verbal text can alter the translator's interpretation of the verbal text contents. Logically, different sets of illustrations would affect the interpretation in different ways.

*Images might affect one translator's solutions differently than they would affect another's.*

Since a viewer's interpretation of an image is affected by the viewer's individual preferences and prior knowledge, one can presume that translators interpret images in their own, slightly differing ways. The way in which a particular image might affect translation will hence depend on the translator.

*Images may consolidate a translator's interpretation of a verbal source text element.* As described earlier in the article, the information conveyed by the visual mode can specify the information conveyed by the verbal mode when their messages are aligned. It might then be possible that the information provided by the image could consolidate the translator's interpretation of a possibly ambiguous element in the verbal text.

*The task of the translator may be affected if visual and verbal information challenge or contradict each other.*

Referring back to the studies conducted by Schnotz and Bannert (2003) and Schnotz and Kürschner (2008) introduced above, it may be assumed that if a verbal text is accompanied by images that do not support the verbal subject matter, the images are likely to have a negative effect on verbal text comprehension. Since the translator's language processing involves verbal text comprehension just like any other reader's, these circumstances may also complicate the translator's efforts to arrive at an adequate translation solution. Lamentably, research on how the combination of visual and verbal modes is comprehended offers no insights into what may happen if the information conveyed by the two modes is contradictory, but one might reasonably assume that these circumstances, too, might complicate the task of the translator.

*Different types of images affect the translation process in different ways.*

The above discussion has been made in the context of technical translation, building on research investigating the interpretation of technical images. Yet it has been suggested that different types of images guide the reading of accompanying verbal text to different degrees: while explanatory technical images are likely to affect the way in which accompanying verbal information is organized and integrated within the cognitive system, images that have been added to verbal text merely to decorate it may not notably guide the cognitive processing of the verbal text (Mayer 1993, 263–267) – consider, for instance, randomly adding an image of a flower bouquet to decorate a love poem. We may therefore assume that different types of images affect the translator's interpretation of the verbal text in different ways; consequently, different types of images affect the translation of verbal text in different ways. This is to say that the fact that a source text includes an image does not automatically mean that the content of the image will affect translation.

In sum, this article has proposed that when translating illustrated technical texts, translation solutions are built on information negotiated from two different modes, even though it has to be emphasized, once more, that readers themselves decide, either consciously or unconsciously, to what extent they process images, and it is therefore highly unlikely that multimodal integration of meaning would happen at every instance of translating segments of illustrated technical texts. The notion has significant implications for translation studies.

Images play an essential role in the overall meaning construction of illustrated texts: they may shape the interpretation of the verbal text and affect the way in which it is translated. The discipline should hence acknowledge images as research objects in their own right. Further, the present discussion can prove to be of importance for neighbouring disciplines as well. If a translated illustrated text segment reflects an instance of how meaning may be extracted from the interaction of verbal and visual information, studying these translation solutions offers one possible tool to exemplify the nature of the semiotic space these two modes create between them. In other words, translation studies has a lot to offer to the study of multimodal meaning making. Yet, a great deal of further research is needed to develop an understanding of the complex issue. Therefore, the article wishes to encourage future research to employ, test, and further develop these ideas.

### Note on contributor

Anne Ketola is a PhD candidate in the School of Language, Translation and Literary Studies at the University of Tampere, Finland, and a member of the MULTI (Multimodality in Translation and Interpreting) research group, investigating the role of multimodality within translation studies. This article is a part of her doctoral thesis in which she examines the interaction of visual and verbal information within the translation of illustrated technical texts. She is also a co-author of the forthcoming monograph *Revoicing Picturebooks* (together with Riitta Oittinen and Melissa Garavini), examining the role of visual information within picturebook translation.

## References

- Anderson, John R. 1995. *Cognitive Psychology and Its Implications*. New York: W. H. Freeman and Company.
- Boeriis, Morten, and Jana Holsanova. 2012. "Tracking visual segmentation. Connecting semiotic and cognitive perspectives." *Visual communication* 11 (3): 259–281. doi: 10.1177/1470357212446408.
- Byrne, Jody. 2012. *Scientific and Technical Translation Explained*. Manchester: St. Jerome.
- Connors, Sean P. 2013. "Weaving multimodal meaning in a graphic novel reading group." *Visual Communication* 12 (1): 27–53. doi: 10.1177/1470357212462812.
- Englund Dimitrova, Birgitta. 2010. "Translation Process." In *Handbook of Translation Studies*, Vol. I, edited by Yves Gambier and Luc van Doorslaer, 406–411. Amsterdam: John Benjamins.
- Eysenck, Michael W., and Mark T. Keane. 2005. *Cognitive Psychology*. Hove: Psychology Press.
- Foulsham, Tom, and Geoffrey Underwood. 2008. "What can saliency models predict about eye movements? Spatial and sequential aspects of fixations during encoding and recognition." *Journal of Vision* 8 (2): 1–17. doi: 10.1167/8.2.6.
- Franco Aixelá, Javier. 2004. "The Study of Technical and Scientific Translation: An Examination of its Historical Development." *Jostrans, The Journal of Specialised Translation* 1, 29–49. [http://www.jostrans.org/issue01/art\\_aixela.php](http://www.jostrans.org/issue01/art_aixela.php).
- Gibbons, Alison. 2012. *Multimodality, Cognition, and Experimental Literature*. New York: Routledge.
- Glenberg, Arthur M., and William E. Langston. 1992. "Comprehension of illustrated text: Pictures help to build mental models." *Journal of Memory and Language* 31 (2): 129–151.
- Göpferich, Susanne, Arnt Lykke Jakobsen and Inger M. Mees. 2009. "Introduction." In *Behind the Mind: Methods, Models and Results in Translation Process Research*, edited by Susanne Göpferich, Arnt Lykke Jakobsen and Inger M. Mees, 1–9. Copenhagen: Samfundslitteratur.
- Halliday, Michael A. K. 1978. *Language as a social semiotic: The social interpretation of language and meaning*. London: Edward Arnold.
- Hegarty, Mary, and Marcel A. Just. 1989. "Understanding Machines from Text and Diagrams." In *Knowledge Acquisition from Text and Pictures*, edited by Heinz Mandl and Joel R. Levin, 171–194. Amsterdam: Elsevier.
- Hegarty, Mary, and Marcel A. Just. 1993. "Constructing Mental Models of Machines from Text and Diagrams." *Journal of Memory and Language* 32 (6): 717–742.
- Henderson, John M., James R. Brockmole, Monica S. Castelhano, and Michael Mach. 2007. "Visual Saliency Does Not Account for Eye Movements During Visual Search." In *Eye Movements: A Window on Mind and Brain*, edited by Roger P.G. van Gompel, Martin H. Fischer, Wayne S. Murray and Robin L. Hill, 537–562. Amsterdam: Elsevier.
- Hiippala, Tuomo. 2012. "The Localisation of Advertising Print Media as a Multimodal Process." In *Multimodal Texts from Around the World: Cultural and Linguistic Insights*, edited by Wendy L. Bowcher, 97–122. London: Palgrave.

- Hirvonen, Maija, and Liisa Tiittula. 2010. "A method for analysing multimodal research material: audio description in focus." *MikaEL, Electronic proceedings of the KäTu symposium on translation and interpreting studies* 4: 1–12.  
[http://www.sktl.fi/@Bin/40698/Hirvonen%26Tiittula\\_MikaEL2010.pdf](http://www.sktl.fi/@Bin/40698/Hirvonen%26Tiittula_MikaEL2010.pdf).
- Horton, William. 2005. "Graphics: The Not Quite Universal Language." In *Usability and Internationalization of Information Technology*, edited by Nuray Aykin, 157–187. Mahwah, NJ: Erlbaum.
- Hull, Glynda A., and Mark E. Nelson. 2005. "Locating the Semiotic Power of Multimodality." *Written Communication* 22 (2): 224–261.
- Jakobsen, Arnt Lykke, and Kristian T.H. Jensen. 2008. "Eye movement behaviour across four different types of reading task." In *Looking at Eyes. Eye-Tracking Studies of Reading and Translation Processing*, edited by Susanne Göpferich, Arnt Lykke Jakobsen and Inger Mees, 103–124. Copenhagen: Samfundslitteratur.
- Jensen, Kristian T.H. 2010. "Distribution of Attention Between Source Text and Target Text During Translation." In *Cognitive Explorations of Translation: Eyes, Keys, Taps*, edited by Sharon O'Brien, 215–236. New York: Continuum International.
- Jewitt, Carey. 2009a. "Introduction. Handbook rationale, scope and structure." In *The Routledge Handbook of Multimodal Analysis*, edited by Carey Jewitt, 1–13. London: Routledge.
- Jewitt, Carey. 2009b. "An introduction to multimodality." In *The Routledge Handbook of Multimodal Analysis*, edited by Carey Jewitt, 14–39. London: Routledge.
- Kaindl, Klaus. 2004. "Multimodality in the translation of humour in comics." In *Perspectives on Multimodality*, edited by Eija Ventola, Charles Cassidy and Martin Kaltenbacher, 173–192. Amsterdam: John Benjamins.
- Kaindl, Klaus. 2013. "Multimodality and translation." In *The Routledge Handbook of Translation Studies*, edited by Carmen Millán and Francesca Bartrina, 257–269. London: Routledge.
- Kress, Gunther, and Theo van Leeuwen. 2001. *Multimodal discourse: the modes and media of contemporary communication*. London: Arnold.
- Kress, Gunther, and Theo van Leeuwen. 2006. *Reading Images. The Grammar of Visual Design*. 2nd ed. London: Routledge.
- Lauer, Claire. 2009. "Contending with Terms: 'Multimodal' and 'Multimedia' in the Academic and Public Spheres." *Computers and Composition* 26 (4): 225–239.  
doi:10.1016/j.compcom.2009.09.001.
- Lautenbacher, Olli Philippe. 2012. "From still pictures to moving pictures. Eye tracking text and image." In *Eye tracking in audiovisual translation*, edited by Elisa Perego, 135–155. Rome: Aracne.
- Lemke, Jay. 2002. "Travels in hypermodality." *Visual Communication* 1 (3): 299–325.  
doi:10.1177/147035720200100303.
- Mayer, Richard E. 1989. "Models for understanding." *Review of educational research* 59 (1): 43–64.
- Mayer, Richard E. 1993. "Illustrations That Instruct." In *Advances in Instructional Psychology*, Vol. 4, edited by Robert Glaser, 253–284. Mahwah, NJ: Erlbaum.

- Mayer, Richard E. 2002. "Using Illustrations to Promote Constructivist Learning from Science Text." In *The Psychology of Science Text Comprehension*, edited by José Otero, José A. Léon and Arthur C. Graesser, 333–356. Mahwah, NJ: Erlbaum.
- Mayer, Richard E. 2005. "Cognitive Theory of Multimedia Learning." In *Cambridge Handbook of Multimedia Learning*, edited by Richard E. Mayer, 31–48. New York: Cambridge University Press.
- Mayer, Richard E., and Joan K. Gallini. 1990. "When Is an Illustration Worth Ten Thousand Words?" *Journal of Educational Psychology* 82 (4): 715–726.
- Mills, Mark, Andrew Hollingworth, Stefan Van der Stigchel, Lesa Hoffman and Michael D. Dodd. 2011. "Examining the influence of task set on eye movements and fixations." *Journal of Vision* 11 (8): 1–15. doi:10.1167/11.8.17.
- Muñoz Martin, Ricardo. 2010. "On paradigms and cognitive translatology." In *Translation and Cognition*, edited by Gregory M. Shreve and Erik Angelone, 169–187. Amsterdam: John Benjamins.
- Nord, Christiane. 1991. *Text Analysis in Translation. Theory, Methodology, and Didactic Application of a Model for Translation-Oriented Text Analysis*. Translated from German by Christiane Nord and Penelope Sparrow. Amsterdam: Rodopi.
- O'Brien, Sharon. 2006. "Eye-tracking and translation memory matches." *Perspectives: Studies in Translatology* 14 (3): 185–205.
- O'Brien, Sharon. 2010. "Introduction." In *Cognitive Explorations of Translation: Eyes, Keys, Taps*, edited by Sharon O'Brien, 1–14. New York: Continuum International.
- O'Halloran, Kay L. 2008. "Systemic functional-multimodal discourse analysis (SF-MDA): constructing ideational meaning using language and visual imagery." *Visual Communication* 7 (4): 443–475. doi:10.1177/1470357208096210.
- Oittinen, Riitta. 1990. "The Dialogic Relation between Text and Illustration: A Translatological View." *TEXTconTEXT* 5: 40–52.
- Oittinen, Riitta. 2000. *Translating for Children*. New York: Garland Publishing.
- Oittinen, Riitta. 2008. "Audiences and Influences: Multisensory Translations of Picturebooks." In *Whose story? Translating the Verbal and the Visual in Literature for Young Readers*, edited by María González Davies, and Riitta Oittinen, 5–18. Newcastle upon Tyne: Cambridge Scholar.
- O'Sullivan, Carol. 2013. "Multimodality as challenge and resource for translation." *Jostrans, The Journal of Specialised Translation* 20: 2–14.  
[http://www.jostrans.org/issue20/art\\_osullivan.php](http://www.jostrans.org/issue20/art_osullivan.php).
- Parkin, Alan J. 2014. *Essential Cognitive Psychology*. Hove: Psychology Press.
- Prieto Velasco, Juan Antonio. 2009. *Traducción e imagen: la información visual en textos especializados*. Granada: Ediciones Tragacanto.
- Prieto Velasco, Juan Antonio. 2012. "A corpus-based approach to the multimodal analysis of specialized knowledge." *Language Resources and Evaluation* 47 (2): 399–423. doi:10.1007/s10579-012-9204-1.
- Reimann, Peter. 2003. "Multimedia learning: beyond modality." *Learning and Instruction* 13 (2): 245–252.
- Remael, Aline, and Josélia Neves. 2007. "Introduction. A tool for social integration? Audiovisual translation from different angles." *Linguistica Antverpiensia* 6: 11–22.

- Ruiz, Carmen, Natalia Paredes, Pedro Macizo, and María Teresa Bajo. 2008. "Activation of lexical and syntactic target language properties in translation." *Acta Psychologica* 128: 490–500. doi:10.1016/j.actpsy.2007.08.004.
- Sager, Juan C. 1994. *Language Engineering and Translation: Consequences of Automation*. Amsterdam: John Benjamins.
- Schnotz, Wolfgang. 2005. "An Integrated Model of Text and Picture Comprehension." In *Cambridge Handbook of Multimedia Learning*, edited by Richard E. Mayer, 49–69. New York: Cambridge University Press.
- Schnotz, Wolfgang, and Maria Bannert. 2003. "Construction and interference in learning from multiple representation." *Learning and Instruction* 13 (2): 141–156. doi:10.1016/S0959-4752(02)00017-8.
- Schnotz, Wolfgang, and Christian Kürschner. 2008. "External and internal representations in the acquisition and use of knowledge: visualization effects on mental model construction." *Instructional Science* 36 (3): 175–190. doi:10.1007/s11251-007-9029-2.
- Schnotz, Wolfgang, and Richard Lowe. 2003. "Introduction: External and internal representations in multimedia learning." *Learning and Instruction* 13 (2): 117–123. doi: 10.1016/S0959-4752(02)00015-4.
- Schnotz, Wolfgang, Maria Bannert, and Tina Seufert. 2002. "Toward an Integrative View of Text and Picture Comprehension: Visualization Effects on the Construction of Mental Models." In *The Psychology of Science Text Comprehension*, edited by José Otero, José A. Léon and Arthur C. Graesser, 385–416. Mahwah, NJ: Erlbaum.
- Shreve, Gregory M., and Erik Angelone. 2010. "Translation and cognition. Recent developments." In *Translation and Cognition*, edited by Gregory M. Shreve and Erik Angelone, 1–13. Amsterdam: John Benjamins.
- Sorden, Stephen D. 2013. "The Cognitive Theory of Multimedia Learning." In *The Handbook of Educational Theories*, edited by Beverly J. Irby, Genevieve Brown, Rafael Lara-Alecio and Shirley Jackson, 155–168. Charlotte, NC: Information Age Publishing.
- Tercedor, Maribel, Esperanza Alarcón-Navío, Juan A. Prieto-Velasco, and Clara I. López-Rodríguez. 2009. "Images as part of technical translation courses: implications and applications." *Jostrans, The Journal of Specialised Translation* 11: 143–168. [http://www.jostrans.org/issue11/art\\_tercedoretal.php](http://www.jostrans.org/issue11/art_tercedoretal.php).
- Tercedor-Sánchez, María Isabel, and Francisco Abadía-Molina. 2005. "The Role of Images in the Translation of Technical and Scientific Texts." *Meta: Translators' Journal* 50 (4). doi: 10.7202/019857ar.
- Youngs, Suzette, and Frank Serafini. 2013. "Discussing Picturebooks Across Perceptual, Structural and Ideological Perspectives." *Journal of Language and Literacy Education* 9 (1): 189–200.

# **TRANSLATION DIARIES OF AN ILLUSTRATED TECHNICAL TEXT**

**Translation students' conceptions of  
word–image interaction**

Anne Ketola

*University of Tampere, Finland*

This study set out to characterize the qualitatively different ways in which a group of master's level translation students conceptualized the interaction of verbal and visual information during the translation of an illustrated technical text. The research approach chosen for this purpose was phenomenography, which aims to identify the differences in the ways people conceptualize various phenomena. The data of the study consisted of translation diaries written by the students.

The study identified two qualitatively distinct main categories of conceptualizing the interaction of verbal and visual information: either conceptualizing the combination of modes as an entity to be perceived as a whole, or conceptualizing the modes as competing sources of information. It was concluded that the students conceptualized the images as an important part of the source text, capable of amplifying, specifying and even annulling verbal information.

**Keywords.** Multimodality, Technical texts, Illustrated texts, Translation students, Translation diaries, Phenomenography.

Translating today often involves engaging with multimodal material (Hirvonen and Tiittula, 2010, p. 1). A multimodal source text conveys meaning through the interaction of modes—written language, spoken language, images, etc. (e.g., Gibbons, 2012, p. 8). The subjects of the present article—illustrated technical source texts—create their message in the interaction of words and images, here



**CONNEXIONS • INTERNATIONAL PROFESSIONAL COMMUNICATION JOURNAL**

2015, 3(2), 13–40

ISSN 2325-6044



referred to as the *verbal source text* and the *visual source text*. The definition of an illustrated technical text adopted in the article builds on Byrne's 2012 (pp. 26–28) description of technical texts, which includes all task-oriented documents that seek to help a particular audience understand how to do something or how something works. By an *illustrated technical text*, the article hence refers to an informative, instructive text that explains how something works by both verbal and visual means. Illustrated technical texts could well be the most common type of illustrated texts being translated today; technical texts constitute a significant share of all translated material (Kingscott, 2002, p. 247, Byrne, 2012, p. 6) and, as remarked by Byrne (2012, p. 26, 54) and Tercedor et al. (2009, p. 143), different types of images are an integral feature of technical documents. Yet, research into the translation of illustrated technical texts so far has been rather scarce, focusing mainly on providing criteria for choosing appropriate images in technical and scientific texts (Tercedor-Sánchez and Abadía-Molina, 2005) and technically-oriented terminological databases (Prieto Velasco, 2009, 2012).

Verbal information in a multimodal text is always interpreted in relation to all modes present (Jewitt, 2009, p. 2), whether consciously or unconsciously. This article sets out to examine how verbal information is interpreted in relation to visual information when translating an illustrated technical text. Naturally, a phenomenon such as this one may be examined from various perspectives. One possible perspective to examining a phenomenon is to describe the distinctively different ways in which people experience and conceptualize it. This article examines the interaction of visual and verbal information within an illustrated technical source text as conceptualized by a group of master's level translation students. The research approach adopted for this purpose is phenomenography, which aims to map the different ways in which a phenomenon or an aspect of the world may be experienced, conceptualized, perceived, and understood by different people (Marton, 1994, p. 4424; 1988, p. 144). Phenomenography sets out to examine *how* people comment on a particular phenomenon, and its aim is to describe the *variation* in the ways of conceptualizing the phenomenon.

The data of the study consist of translation diaries—reports on the problems encountered, the strategies employed to solve them, and so on—written

about the translation process of an illustrated technical text during a technical translation course from English to Finnish. The students were not specifically instructed to comment on the images or their interaction with the verbal text in their diaries. For this reason, prior to analyzing *how* the students comment on the interaction of verbal and visual information, the analysts must determine *if* the students comment on the issue—do they inspect the images and do they pay attention to their interaction with the verbal text? If the students were to regard the images, for instance, as merely decorative elements, it is unlikely that they would comment on them when discussing their translation strategies. The research questions that the article sets out to answer may hence be formulated as follows: “Are translation students aware of the interaction of verbal and visual information in illustrated technical texts?” and “How do translation students conceptualize the interaction of verbal and visual information within the translation of an illustrated technical text?”

## **Background of the Research Project**

The study presented in this article is a part of a research project investigating how an illustrated technical text is processed in translation. The research data of the project consists of the translations of an illustrated technical text made by translation students as well as the translation diaries examined in this article. The data were produced during a technical translation course at the University of Tampere, Finland. The research subjects were a group of eight master's level translation students who all spoke Finnish as their native language. All of the students had received a bachelor's degree in English translation at the University of Tampere, having therefore completed both theoretical as well as practical translation courses—both from and into English. The students were given one week to finish the translation assignment. The use of dictionaries and other reference resources was allowed. After the students had completed the translation task, a group interview was conducted in which the students were informed of the aim of the research and where they had the chance to comment on the task. The research project was submitted for evaluation by the university's ethics committee

for research involving human subjects and it was exempted from the need for a review. All eight students gave written consent to participate in the study.

The source text for the translation assignment presented the illustrated operating principles of two different types of wet magnetic separation devices used in the mining industry for ore beneficiation. The source text, including its illustrations, was produced specifically for the purposes of this research with the help of the staff of the Geological Survey of Finland. It was written in English and proofread by an expert at the Western Australian School of Mines. The source text consisted of just over 500 words and two large colored images, and I consider it to be a representative example of an illustrated technical text as defined above.

A broad range of research has established that, when reading an illustrated text, readers process both words and images, and form their interpretation of the multimodal text based on both verbal and visual information (e.g., Connors, 2013; Hegarty and Just 1993; Mayer, 2002; 2005; Schnotz and Bannert, 2003; Schnotz and Kürschner, 2008; Wasylewsky and Tapajna, 2001; Youngs and Serafini, 2013). The information provided by the two modes may hence be deeply intertwined in the reader's mind. When designing the source text for the study, I considered it possible that if the messages conveyed by the two modes were perfectly symmetrical—in other words, they expressed precisely corresponding information—then the students might not be able to distinguish which parts of their interpretation consisted of information derived verbally and which visually. Hence, they could be inclined to underestimate the importance of the visual information within the process even if they had inspected the images with great care; after all, it is words that they undoubtedly acknowledge reading, and it is words that they produce.

For this reason, the relationship between the words and the images in the source text was modified in certain parts of the source text so that the information conveyed by the two modes was, in one way or another, *asymmetrical*. For instance, in one section of the source text, visual information was deleted from the image: the verbal text accurately described a particular part of the operating process—tailings or nonmagnetic particles exiting the separator—but the

corresponding information could not be found in the image. On two occasions, the visual and verbal texts were modified so that the information provided by the two was straightforwardly contradictory: the verbal text expressed that the device was submerged under water, while only the bottom-most part of the device was under water in the image, and the verbal text expressed that a certain part of the device was located in the *upper* part of the device while, according to the image, the part was in the *lower* part of the device. Further, the shape of a part of the device called *launder* was modified: while the term typically refers to a trough or a long, narrow container, it was presented in the image as nearly square-shaped. The rationale behind these modifications was that the asymmetry of information might make it easier to distinguish which mode the translation student considered to be of more relevance during translation.

In a pre-translation questionnaire, half of the students estimated their background knowledge of ore beneficiation as “none” and half as “very little.” Even though these estimations are rather subjective, the students’ level of background knowledge of the subject matter is an important factor to consider: previous research within educational psychology investigating university students’ reading comprehension of illustrated scientific texts has established that students with low prior knowledge of the subject matter at hand observe the illustrations in more detail than students with high prior knowledge (Mayer and Gallini, 1990). The fact that none of the students reported being considerably better acquainted with the subject matter than the rest of the group adds to the comparability of the data.

The present article concentrates on examining a relatively small number of translation students. The study does not intend to generalize its observations to all translation students and, even less so, to professional translators who may generally be expected to employ translation strategies differing from those employed by translation students (e.g., Götferich, 2010). It should be emphasized that in the discussion of the results of the analysis, the term “translator” as opposed to “translation student” is employed; in that context, it is used to refer to “the person who translated something.”

## **Translation Diaries as Research Data**

Translation students at the University of Tampere are asked to write translation diaries of a large majority of the translation assignments that they are given during the study program. They are generally instructed to comment on the communicative situation—commissioner of the translation, the target audience, and so on—and on the source text itself—its style, text type, and subject area. They are also asked to describe the translation process—macro and micro level strategies—and to specify the sources consulted, including dictionaries and parallel texts—texts originally written in the target language with a purpose similar to that of the source text. Finally, they are instructed to reflect on the quality of their own work and to assess what they learned during the assignment (UTA käänöskomenttiohje). These translation diaries could be described as *semistructured* as instructions for writing the diaries are generally provided in the first year of the study program and, after the students are accustomed to writing them, the instructions are not regularly reinforced.

This translation diary procedure is highly similar to the Integrated Problem and Decision Reporting (IPDR) procedure introduced by Gile (2004, p. 15): both set “a systematic requirement for written introspective reporting by students whenever they hand in a translation assignment.” In both forms of reporting, students discuss the problems they encountered during the translation process, the steps they took to solve them, and the rationale behind their decisions. Students also introduce the sources and references they consulted during the task. The reports are collected in a written form with no strict reporting format or structure except for the initial instructions from the instructors (Gile, 2004, pp. 3–4). Introspective reporting benefits the students and the instructors alike. Gile describes that writing reports increases the students’ awareness of their translation process: it emphasizes that translation is a demanding operation requiring intense decision-making, and it encourages the students to devote more effort to their work. It also provides the instructors with a better view to what the students are doing and what they find particularly

difficult, as well as how to interpret their translations and to identify their strategies (pp. 4–9).

Gile, too, remarks that apart from having a didactic function, these introspective reports may also be used as data for research (pp. 8–9). Translation process research has employed other introspection methods as well, including think-aloud methods and retrospection (e.g., Tirkkonen-Condit and Jääskeläinen, 2000). Gile (2004, pp. 8–9) emphasizes that IPDR offers “no revolutionary way of accessing information not available through other methods,” but that compared to other introspection methods, it does have its benefits: the data is easily gathered and available to the researcher in a directly readable form. Further, IPDR does not include distraction between translating and verbalizing one’s thoughts at the same time, nor does it require the students to work under strict time limitations or with a particular software. Gile (p. 10) concludes that the main limitation of IPDR as a data collection method lies in the unpredictable, possibly noncomprehensive nature of the data, and suggests that more complete reports could perhaps be produced with the help of a more structured set of questions posed for the translators. Göpferich and Jääskeläinen (2009, p. 172) make the same observation, stating that the contents of the reports depend entirely on what the translators themselves regard as relevant. Yet, the justification for the use of the data collection method in this particular study lies in that very notion: one of the aims of the study was to elucidate whether the translation students indeed regard visual information as relevant enough to analyse in their diaries. In addition, as discussed in the following section, it would be highly questionable to perform a phenomenographic analysis on translation diaries written in response to a structured set of questions.

## **Research Approach and Process**

The research approach adopted in the analysis of the translation diaries is phenomenography, which sets out to map the different ways in which various aspects of the world are experienced, conceptualized, perceived, and understood by different people (Marton, 1994, p. 4424; 1988, p. 144). Phenomenography was

developed and has mainly been applied within an educational context. Yet, its aim transcends the educational context as the approach sets out to identify similarities and differences in the way we experience and comprehend phenomena in the world around us (Marton, 1994, p. 4429).

Rather than a single method of analysis, phenomenography is an integral perspective on research. It is underpinned by the adoption of a second-order perspective, which refers to focusing on how things appear to people instead of focusing on how things “really are” in the world. In other words, within the so-called first-order perspective, research aims to make statements about the world (Marton, 1981, p. 178) and could pose a question such as “How do words and images interact in illustrated texts?” Within the second-order perspective adopted in phenomenography, on the other hand, research aims to make statements about people’s conceptions of the world (*ibid.*), posing questions such as “How do translators themselves experience and conceptualize word–image interaction in illustrated texts?” In order to fully understand the phenomenon of word–image interaction within translation, observations made from the first-order perspective may be complemented by those made from the second-order perspective.

## **Phenomenographic Analysis and Issues of Data Collection**

The data used in phenomenographic research is collected from a group of people individually reflecting on their experience of a phenomenon (Reed, 2006, p. 5). The researcher goes through the data and looks for expressions—direct quotes—that refer to experiencing the phenomenon under study in a certain way. Based on their similarities and differences, these experiences are arranged into groups referred to as *categories of description* (Marton, 1988, p. 145). In other words, the categories are not selected in advance but emerge from the data as expressions are brought together and compared. As the same individual may express more than one way of conceptualizing the phenomenon when in different situations, the individual—in this case, the individual translator—is not the unit of analysis (Marton and Pong, 2005, p. 346). Instead, the categories of description are

arrived at by “separating forms of thought both from the thinking and the thinker” (Marton, 1981, p. 196).

Since the categories are the result of reflections about the same phenomenon, they are meaningfully related to each other. Further, as some ways of experiencing a phenomenon may be more comprehensive than others in relation to a particular criterion, it may be possible to establish a *hierarchy* between the categories of description (Marton, 1994, p. 4426). The main aim of phenomenographic analysis is to identify the relationships or the structure between the categories. The structured set of categories of description is called *the outcome space* of the phenomenon in question, and it constitutes the main outcome of phenomenographic research (Marton, 1994, p. 4424). Since phenomenographic research in general—this study being no exception—analyzes a relatively small number of research subjects, the outcome space may never be claimed to be exhaustive. The goal is simply that the outcome space is complete in the sense that nothing has been left out of the collective experience of the particular group (Marton and Booth, 1997, p. 125).

The dominant method of data collection in phenomenographic research has traditionally been the individual interview, but some studies have also employed group interviews, children’s drawings, written responses, and historical documents (Marton, 1994, p. 4427), as well as unstructured learning diaries (Prinsloo, Slade, and Galpin, 2011) as research data. As explained above, the data used in this study consists of semistructured translation diaries written by a group of translation students. This type of data has both advantages and disadvantages as research data for phenomenographic inquiry. An obvious disadvantage is that the method of data collection does not offer a chance to ask the students for further clarifications of their reflections. A clear advantage, on the other hand, is that the researcher cannot guide the students’ reflections by—consciously or unconsciously—imposing one’s own presuppositions of the phenomenon in the form of follow-up questions. Marton (1988, p. 154) emphasizes that any guiding questions used within the process of phenomenographic data collection should be as open-ended as possible, allowing for the subjects to “choose the dimensions of the question they want to answer.” The dimensions of the answer reflect what the

subject holds relevant and are therefore informative in themselves. Since the translation students of the study were not instructed to comment on the images, it may be concluded that all comments regarding visual information and word-image interaction were made because the students regarded them as relevant issues to discuss.

## **Procedure of Analysis Followed in the Study**

Exact rules of procedure for undertaking phenomenographic analysis cannot be specified: as Marton (1988, p. 154) concludes, finding out the ways in which a phenomenon is conceptualized by different individuals “takes some discovery” for which no algorithms can be provided. Yet, a general way of proceeding can be described; the procedure followed in this study is based on the general guidelines provided by Marton (1988, pp. 154–155), Marton (1994, p. 4428), and Larsson and Holmström (2007, p. 57). Even though the number of steps and the use of some terminological choices differ in these guidelines, the key elements provided remain the same for them all.

The first stage of analysis included reading through the data various times. The data consisted of eight translation diaries ranging from 400 to 900 words. I observed that seven out of the eight translation students had commented on issues relating to visual information and therefore concluded that the analysis was indeed possible. When reading through the data, my aim was first to gain a tentative understanding of what the students said and then, with each rereading, to obtain a more comprehensive understanding of the data as a whole. In the second stage of the analysis, I began to select parts of the data that answered the question “What do the translation students say about images or the interaction of images and words?” This guiding question was deliberately formulated as a rather open one in order to avoid misjudging what the students regarded as relevant. The selected parts ranged from individual phrases to parts of longer reflection. At the end of this stage, I had gathered a collection of quotes—the “data pool” of my study—consisting of 53 quotations. The translation diary, which included no

references in regard to the visual, was consequently left outside of the analysis at this point.

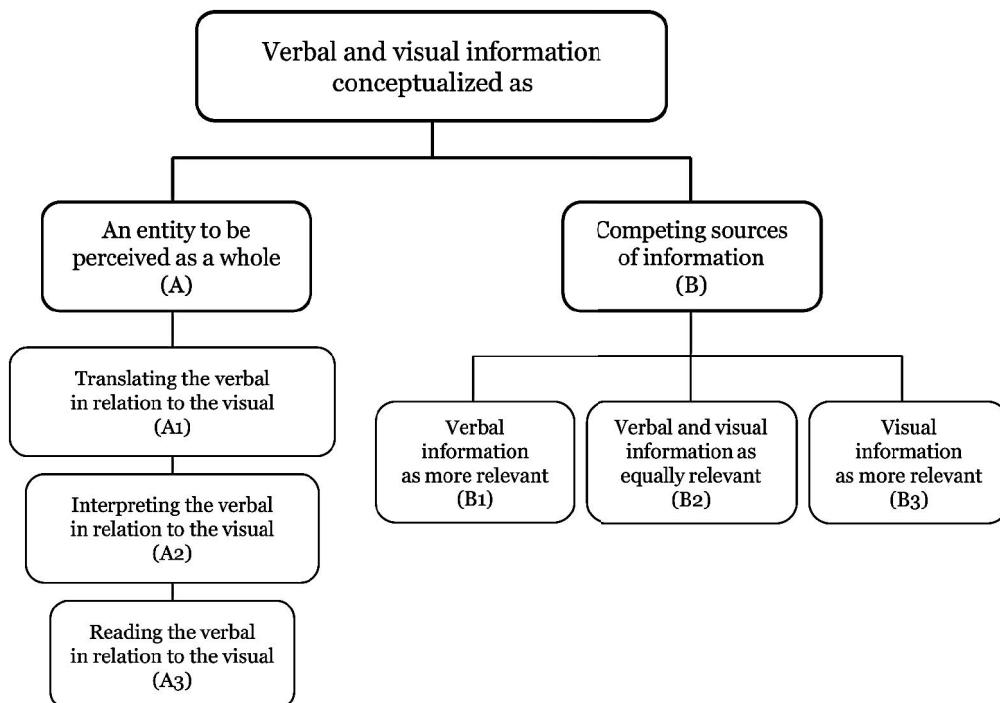
In the third stage of the analysis, my attention shifted from the individual students to the meaning embedded in the data pool as a whole. In other words, at this stage of phenomenographic analysis, forms of thought are conclusively separated from the thinker; beyond this point, the analysis no longer specifies which research subject each quotation belongs to. At this stage, my aim was to identify the different ways of understanding the phenomenon, the guiding question being “What are the different ways in which the translation students conceptualize the interaction of verbal and visual information?” I read through the quotes again and began to perceive some general themes that ran across the pool. The quotations were arranged into (tentative) groups based on their similarities and differences. Eventually, this led to establishing clearer borders between the groups. It was then possible to detect and determine the distinguishing features of each group. These groups now formed the categories of description of my study. Structuring the outcome space, the final stage of analysis proceeded hand in hand with the previous stage. This outcome space constitutes the final result of the analysis; it represents a contrastive comparison of the different ways of conceptualizing the interaction of verbal and visual information in the particular source text by the particular group of translation students.

## **Results of the Phenomenographic Analysis**

The results of the analysis are presented in Figure 1 on page 72 as a set of categories of description in an outcome space, symbolizing the range of the different ways in which the phenomenon under investigation was conceptualized collectively. To demonstrate that the categories of description are indeed supported by empirical data, the categories are presented with example quotes from the data (Francis, 1996, p. 44; Åkerlind, 2005, pp. 331–332). The quotations have been translated from Finnish into English. Since the quotations have been removed from their original contexts, I added contextual information in brackets when considered necessary. Even though elsewhere in the article *a text* consisting

**Figure 1**

The outcome space of translation students' ways of conceptualizing the interaction of visual and verbal information in the source text.



of verbal information only is referred to as *a verbal text*, the students, quite naturally, refer to verbal text simply as “text” in the quotations.

The analysis identified two qualitatively distinct main categories of description representing the translation students' ways of conceptualizing the interaction of verbal and visual information in the source text:

- A. Conceptualizing the verbal and the visual as an entity to be perceived as a whole
- B. Conceptualizing the verbal and the visual as competing sources of information

In other words, the first main category represents experiences related to perceiving the combination of word and image as an inseparable whole, and the second represents experiences related to consciously dissecting the combination of verbal and visual information and evaluating the two modes as separate sources of information. Both main categories include subcategories; the subcategories in group A take the form of a linear hierarchy of inclusiveness—the lower categories are logically included in the higher ones—whereas the subcategories in group B are qualitatively exclusive of each other.

### **Main Category A: Conceptualizing the Verbal and the Visual as an Entity to be Perceived as a Whole**

In the first main category, the combination of word and image was conceptualized as an entity to be read, interpreted and translated as a whole. Characteristic of these experiences was that the information conveyed by the two modes was combined into a single message, the integrality of which was not questioned. This was reflected particularly well in experiences related to instances of asymmetry of information. The translators expressed three hierarchically related aspects to conceptualizing the multimodal source text as a whole. The category is hence divided into three subcategories, hierarchically linked based on their inclusiveness and relative completeness: A3 at the bottom of the hierarchy is the most basic category and A1 at the top is the most comprehensive, logically subsuming both lower subcategories of A2 and A1.

**Reading the verbal in relation to the visual (A3).** This subcategory concentrates on describing how the verbal and the visual are read side by side and how the attention of the translator alternates between the visual and the verbal in different stages of translation. The focus of the subcategory is hence on the external representations: the actual verbal source text and the illustrations. Characteristic of these experiences is that the image is conceptualized as an integral part of the multimodal source text: it is involved within source text inspection in various stages of translation. An example of the source text

comprehension stage is illustrated in the quotation below in which the translator describes stopping to check the image after reading small segments of the verbal source text:

In other words, I read them [*words and images*] side by side; in each part, I stopped to see where it [*the slurry*] really goes.

In the following quotation, the translator reports having spotted an instance of asymmetry between the verbal and visual texts in the revision stage of the translation process:

I didn't notice it until four hours before the deadline and then I sort of panicked.

The quotation illustrates how the revision stage, too, has included switching attention between the two modes; comparing the verbal text (either the source or the target or both) with the image.

**Interpreting the verbal in relation to the visual (A2).** Relatively more complete than the previous subcategory, this subcategory represents experiences of interpreting and comprehending the multimodal source text as a whole instead of merely switching attention between the two modes. In other words, the focus of the category is on interpretation and forming an internal or mental representation of the two external representations. This subcategory logically subsumes the lower subcategory introduced above: interpreting the two modes in relation to each other presupposes that the translator has inspected them both. Many of these experiences are related to resorting to images for confirmation of acquired verbal text contents—even though, admittedly, these experiences only describe the students' own subjective understanding of their comprehension process.

The images acted as elements supporting and confirming my text comprehension.

Images helped me process and comprehend the text.

Interestingly, these experiences were not limited to the images of the source text at hand but were also related to other visual information the translators had employed within the translation process. These images could be described as *visual parallel texts*—images produced for a similar purpose to that of the images of the source text:

Google image search was extremely helpful in assuring I had understood certain parts of the equipment somewhat correctly.

The asymmetry of information between the modes of the source text was experienced as hindering source text comprehension. The following quotations are related to the translators reflecting on the part of the source text where information (exit route of the tailings) was missing from the image:

I understand the operating principle of the equipment rather well, even though I do not understand where the tailings go in the counter current separator.

I still [*after completing the assignment*] do not understand where they [*tailings*] go.

These experiences reflect a clear effort to interpret the multimodal source text as a whole: visual and verbal information are mapped onto each other to a degree that not being able to confirm the verbally acquired information from the image leaves the translator with an impression of not having understood the verbal text at all. Despite the obvious asymmetry of information, the truth value of neither mode is questioned.

**Translating the verbal in relation to the visual (A1).** The most comprehensive subcategory, this subcategory focuses on translating the internal representation formed from the external representations; in other words, the interaction of visual and verbal information. It is highest up in the hierarchy because translation presupposes interpretation. Even though translation can be described as verbal activity in the sense that it is verbal text that the translator produces, these experiences clearly reflect how images define what the most suitable or “the

“correct” translation solution is. Translation solutions are based on or justified with visual information, as exemplified by the following quotation:

Thanks to the images, I had the courage to deviate from the source text more radically. I was more certain that I had understood the text correctly, and was therefore more confident about the suitability of the deviating translation solutions to each given context.

The following two quotations are excellent examples of how the image may specify the meaning of a particular element in the verbal source text and how the image defines what the correct translation solution is. In the first, the translator compares a list of possible translation choices to the information provided by the image, and in the second, the suitability of the translation solution is assessed in relation to visual information:

Judging by the image, the part in question is not a funnel, a gearbox, an access door box or a charge pocket.

“A container” is an easy choice since it is so generic it can be of any shape at all.

Some of these experiences, too, were related to visual parallel texts. The following quotation illustrates how a translation solution is based on visual information found in other images:

After I checked images of similar devices online, I decided to change it [*the translation solution*].

In the following quotation, the translator reflects on the translation of the word “feedbox,” and reports having found a Finnish equivalent for a *similar looking part* labeled in an image describing the production process of biofuel. The external similarity—and, perhaps, the assumed functional similarity—of the parts in the two images is considered to ratify the use of the equivalent in the translation:

I found an image of a similar feedbox in a text describing the production of biofuel. I believe this validates my choice.

All in all, the experiences in main category A indicate a conscious effort to map visual and verbal information onto each other and to apprehend each mode in terms of the other. This way of conceptualizing the interaction of modes did not include questioning the accuracy of either one, even when there was a clear asymmetry of information between the two.

### **Main Category B: Conceptualizing the Verbal and the Visual as Competing Sources of Information**

Category B represents experiences of consciously dissecting the multimodal source text and conceptualizing word and image as separate sources of information. These experiences were related to evaluating the verbal and the visual modes as sources of information, or assessing their usefulness and trustworthiness in relation to each other. Most—but, interestingly, not all—of the experiences in this category were related to the instances of asymmetry of information between the modes. Some of the experiences in the category—in subcategory B3, to be precise—also included examples of the translator assessing the usefulness and trustworthiness of the modes on a more general level: in other words, when the multimodal source text did not directly prompt a reason to do so. Three qualitatively exclusive subcategories were detected, representing the variation in the ways of conceptualizing the two modes as sources of information in relation to each other.

**Conceptualizing verbal information as more relevant (B1).** In this subcategory, visual information was experienced primarily as subsidiary to verbal information. Characteristic of these experiences was that the usefulness of the image was questioned when the two modes provided asymmetrical information:

The second image was not as informative as the text.

These experiences also reflect discrediting the contents of the visual source text when the two modes were contradictory to each other.

I trusted the contents of the text completely and therefore ignored the inadequacy of the image.

In one part, the information in the source text differed from that claimed by the image.

**Conceptualizing verbal and visual information as equally relevant (B2).** This subcategory supports a view of considering word and image as sources of information with an equal status, not explicitly identifying either one as more trustworthy than the other.

There might be a mistake in the paragraph because according to the image, the tailings are collected in the lower part of the machine, and not in the upper part as expressed in the text.

In the above quotation, the asymmetry of information is reported in a somewhat neutral manner: the translator states that one mode expresses this piece of information in one way, and the other mode in another way. The translator does not directly propose that the verbal text is mistaken, but that there *might* be a mistake. This subcategory did not emerge from the data as a prominent one; in fact, the quotation above was the only one representing a view of not directly disregarding either mode. All other quotations referring to conceptualizing the issue reflected a considerably more explicit stance towards privileging either verbal or visual information.

**Conceptualizing visual information as more relevant (B3).** This subcategory was related to experiencing verbal information as subsidiary to visual. Visual information was conceptualized as both amplifying the verbal, in other words, providing information not available from the verbal text, as well as annulling or discrediting the contents of the verbal text when the two modes were contradictory to each other. Out of the three subcategories in category B, B3 emerged from the data as the most prominent one. Unlike the subcategories of B1

and B2, these experiences were not limited to discussing instances of asymmetry of information between the modes but were also made on a more general level; reflection on the visual mode being more informative than the verbal was often related to the comprehension of the subject matter, in other words, the operating principle of devices the source text describes:

The image gave me a better overall understanding of the subject, revealed the shape of the parts and the direction of the flow of the substances etc.

Compared to subcategory A2, in which the image was conceptualized as affirming the acquired verbal text contents, the image is here conceptualized as helping in source text comprehension by offering information not retrievable from the verbal text. In fact, visual information was conceptualized as so intrinsic to the comprehension of the operating principle of the device that the translators reported having imagined it in the parts where visual information was missing from the image:

I looked at the image to assess where they [*particles missing from the image*] actually go.

In this subcategory, the experiences related to conceptualizing the contradiction of information between the two modes reflected conceptualizing the verbal element as mistaken. This was, in fact, the most common way of commenting on the instances of contradiction.

The contradiction between the text and the image made me question the accuracy/incorrectness of the text.

...the part which talked about the location of the tailings launders, claiming it to be in the upper part of the equipment, while the image clearly shows that it is one of the bottommost parts of the equipment.

Another part indicative of the somewhat deficient connection between image and text is the part which says that the drum is submerged in the tank, which I

understand to mean that the drum is entirely under water – yet, this is not the case.

Some of these experiences were also justified with visual parallel texts:

In the image (or in other images I saw) it [*part of the equipment*] is not in the upper part.

Interestingly, even though these experiences reflected a clear tendency to identify the verbal component as mistaken, most of these translators also mentioned that they did not, in fact, change the “mistaken” verbal information in their translations: even though they were convinced that the verbal text was wrong, they reported that they did not feel they had the authority to change it. All in all, main category B suggests that the combination of verbal and visual information may be approached with strategies reflecting varying attitudes towards the importance and relevance of both modes.

The group interview conducted a week after the translation assignments had been handed in shed some further light on the translators’ thought processes. As mentioned above, one of the students did not discuss the images at all in the translation diary. Yet, in the group interview, this particular student reported having studied the images in great detail—however, this had been done even before reading the verbal source text for the first time. In the student’s own words, the student had examined the images first to gain a general idea of the subject matter and had then moved on to read and translate the verbal text, disregarding the images. This illustrates that visual information may be approached with different reading strategies. Further, several students discussed the part of the source text in which visual information had been deleted and mentioned that it had not occurred to them that the image could be faulty in any way—they had simply thought they were “too stupid” to understand the process correctly. Further, one of the students reported having thought that the asymmetry of information could well have been due to difficulties in printing the colors correctly—that, for some reason, the color blue (the color of the deleted

information) had not printed correctly. These notions are discussed in the following section in relation to the results of the phenomenographic analysis.

## **Conclusions**

This study set out to elucidate if a group of translation students were aware of the interaction of verbal and visual information in an illustrated technical source text as well as to characterize the different ways in which the students conceptualized this interaction. Visual information—in particular, instances of asymmetry of information between verbal and visual information—was widely discussed in the translation diaries, which confirmed that most students had indeed inspected the images. Further, in the group interview, the one translator who did not discuss the issue in the translation diary confirmed having inspected the images. In other words, all eight translators inspected the images while reading the source text. Moreover, visual information was considered to be a relevant part of the source text; at times the translators concluded that they esteemed verbal information above the visual, but even such a comment indicates that the translator consciously negotiated the relationship between the modes. In answer to the first research question, it may hence be concluded that the translators were principally aware of the interaction of verbal and visual information, even though one of them only inspected the images before starting the actual translation of the verbal text.

In answer to the second research question, the study identified two distinct main categories of conceptualizing the interaction of verbal and visual information in an illustrated technical source text: either making a conscious effort to read, interpret and translate the combination of the modes as a whole, or separating the two modes of the source text for comparison and evaluation. These categories are not meant to describe individual translators: as often happens in phenomenographic research (Marton, 1994, p. 4428), the same individuals expressed more than one way of conceptualizing the phenomenon when approaching it from different angles.

The experiences in the category of perceiving the combination of words and images as an inseparable whole indicated a conscious effort to map visual and verbal information onto each other. When this was impeded by deleting information from the image, the translators seemed to be left with an impression of not having understood the source text at all. Unlike in the second category, the truth value of neither mode was not questioned. These observations—together with the students' accounts of how they did not doubt the information conveyed by the images, expressed during the group interview—suggest that images are considered to have a relatively high truth value; in other words, that they accurately depict the objects they represent. An obvious failure to do so was explained by doubting one's own sense or blaming it on the reproduction rather than production of the image. When referring directly to translation, the experiences in this category illustrated how translation solutions were negotiated from the interaction of the two modes; the verbal source text evoked various different translation solutions, and the image defined what “the correct” solution was.

In the category of evaluating words and images as competing sources of information, three different stances were identified: translators either considered verbal information as more relevant than the visual, considered them both as equally relevant, or considered visual information as more relevant than the verbal. In other words, the combination of verbal and visual information was approached with strategies reflecting varying attitudes towards the importance and relevance of both modes. It may be reasonably assumed that these underlying attitudes may affect subsequent translation choices: for instance, valuing visual information above the verbal could lead to translation choices based more on visual information than the verbal. Once again, it has to be emphasized that these categories did not characterize individual translators: the same translators adopted different stances towards the two modes in different situations.

In their translation diaries, the students concentrated mainly on discussing the features of the target text and describing their translation process. A few, but not all, discussed the communicative situation and reflected upon the quality of their own work. Yet, the most recurring theme in the translation diaries were the

difficulties produced by the asymmetry of information between the two modes. This suggests that visual and verbal information challenging or contradicting each other disturbed the translation process considerably. The translator who did not employ visual information during the actual translation stage did not discuss these parts of the source text at all in the translation diary, which may be interpreted as implying two things: first, the fact that the rest of the group singled these parts out as difficult to translate was indeed due to lack of coherence between the modes—as opposed to the verbal source text simply being more difficult in these parts. Second, it implies that a verbal text coupled with images may be interpreted and translated differently than the same text without images; the source text without the images is, in a sense, a different source text.

The most significant conclusion that may be drawn from the present study is that the translation students did not merely assign the visual source text to a decorative or contextual role; instead, they conceptualized it as an important source of information, capable of amplifying, specifying, and even annulling elements of the verbal source text. Further, some of the reflection in both categories was made in reference to images other than those constituting the visual source text, referred to in this article as *visual parallel texts*. This reflection indicated that visual parallel texts were considered as useful and reliable sources of information. All in all, the observations made in the article have great consequences for translation studies—a discipline that has traditionally been heavily language-oriented (O’Sullivan, 2013, pp. 2–3; Kaindl, 2013, p. 257). Hence, this article emphasizes the importance of acknowledging images as an object of inquiry in their own right within the discipline, both in translation research as well as translator training.

The observations made in this article from the second-order perspective can contribute to future research on word–image interaction in translation made from the first-order perspective; in other words, studying illustrated source texts or their translations directly. The main limitation of the findings is that the ways of conceptualizing word–image interaction are not an exhaustive description of all the possible ways of conceptualizing the phenomenon. It is likely that an analysis with a different sample of translation students would result in additional

categories of description. An interesting topic for future research would also be to analyse whether professional translators would conceptualize the phenomenon differently than translation students would. Further research could also investigate how translators conceptualize word–image interaction in other illustrated text types, for instance in children’s picture books. ■

## References

- Åkerlind, G. S. (2005). Variation and commonality in phenomenographic research methods. *Higher Education Research and Development*, 24(4), 321–334. doi:10.1080/07294360500284672
- Byrne, J. (2012). *Scientific and technical translation explained*. Manchester: St. Jerome Publishing.
- Connors, S. P. (2013). Weaving multimodal meaning in a graphic novel reading group. *Visual Communication*, 12(1), 27–53. doi:10.1177/1470357212462812.
- Francis, H. (1996). Advancing phenomenography – Questions of method. In G. Dall’Alba & B. Hasselgren (Eds.), *Reflections on phenomenography: Toward a methodology?* (pp. 35–47). Göteborg: Kompendiet.
- Gibbons, A. (2012). *Multimodality, cognition, and experimental literature*. New York: Routledge.
- Gile, D. (2004). Integrated problem and decision reporting as a translator training tool. *Journal of Specialised Translation*, 2, 2–20. Retrieved from [http://www.jostrans.org/issue02/art\\_gile.pdf](http://www.jostrans.org/issue02/art_gile.pdf)
- Göpferich, S. (2010). The Translation of instructive texts from a cognitive perspective: Novices and professionals compared. In S. Göpferich, F. Alves, & I. M. Mees (Eds.), *New approaches in translation process research* (pp. 5–55). Copenhagen: Samfundslitteratur.

Göpferich, S., & Jääskeläinen R. (2009). Process research into the development of translation Competence: Where are we, and where do we need to go? *Across Languages and Cultures*, 10(2), 169–191. doi:10.1556/Acr.10.2009.2.1

Hegarty, M., & Just. M. A. (1993). Constructing mental models of machines from text and diagrams. *Journal of Memory and Language*, 32, 717–742. doi:10.1006/jmla.1993.1036

Hirvonen, M., & Tiittula, L. (2010). A method for analyzing multimodal research material: Audio description in focus. *MikaEL, Electronic proceedings of the KäTu symposium on translation and interpreting studies*, 4, 1–12. Retrieved from [http://www.sktl.fi/@Bin/40698/Hirvonen%26Tiittula\\_MikaEL2010.pdf](http://www.sktl.fi/@Bin/40698/Hirvonen%26Tiittula_MikaEL2010.pdf)

Jewitt, C. (2009). Introduction: Handbook rationale, scope and structure. In C. Jewitt (Ed.), *The Routledge handbook of multimodal analysis* (pp. 1–13). London: Routledge.

Kaindl, K. (2013). Multimodality and translation. In C. Millán & F. Bartrina (Eds.), *The Routledge handbook of translation studies* (pp. 257–269). London: Routledge.

Kingscott, G. (2002). Technical translation and related disciplines. *Perspectives: Studies in translatology*, 10(4), 247–255. doi:10.1080/0907676X.2002.9961449

Larsson, J., & Holmström, I. (2007). Phenomenographic or phenomenological analysis: Does it matter? Examples from a study on anaesthesiologists' work. *International Journal of Qualitative Studies on Health and Well-being*, 2, 55–64. doi:10.3402/qhw.v2i1.4945

Marton, F. (1981). Phenomenography: Describing conceptions of the world around us. *Instructional Science*, 10, 177–200. doi:10.1007/BF00132516

Marton, F. (1988). Phenomenography: A Research approach to investigating different understandings of reality. In R. R. Sherman & R. B. Webb (Eds.), *Qualitative research in education: Focus and methods* (pp. 141–161). London: Falmer Press.

Marton, F. (1994). Phenomenography. In T. Husén & T. N. Postlethwaite (Eds.), *The international encyclopedia of education* (vol. 8), (pp. 4424–4429). Oxford: Pergamon.

- Marton, F., & Booth, S. (1997). *Learning and awareness*. Mahwah: Erlbaum.
- Marton, F., & Pong, W.Y. (2005). On the unit of description in phenomenography. *Higher Education Research & Development*, 24(4), 335–348. doi:10.1080/07294360500284706
- Mayer, R. E. (2002). Using illustrations to promote constructivist learning from science text. In J. Otero, J. A. Léon, & A. C. Graesser (Eds.), *The psychology of science text comprehension* (pp. 333–356). Mahwah: Erlbaum.
- Mayer, R. E. (2005). Cognitive theory of multimedia learning. In R. E. Mayer (Ed.), *Cambridge handbook of multimedia learning* (pp. 31–48). New York: Cambridge University Press.
- Mayer, R. E., & Gallini, J.K. (1990). When is an illustration worth ten thousand words? *Journal of Educational Psychology*, 82(4), 715–726. doi:10.1037/0022-0663.82.4.715
- O’Sullivan, C. (2013). Multimodality as challenge and resource for translation. *The Journal of Specialised Translation*, 20, 2–14.
- Prieto Velasco, J. A. (2009). *Traducción e imagen: la información visual en textos especializados*. Granada: Ediciones Tragatanto.
- Prieto Velasco, J. A. (2012). A corpus-based approach to the multimodal analysis of specialized knowledge. *Language Resources and Evaluation*, 47(2), 399–423. doi:10.1007/s10579-012-9204-1.
- Prinsloo, P., Slade, S., & Galpin, F. (2011). A phenomenographic analysis of student reflections in online learning diaries. *Open Learning: The Journal of Open, Distance and e-Learning*, 26(1), 27–38.
- Reed, B. (2006). Phenomenography as a way to research the understanding by students of technical concepts. *Proceedings of NUTAU 2006*, 1–11.
- Schnotz, W., & Bannert, M. (2003). Construction and interference in learning from multiple representation. *Learning and Instruction*, 13(2), 141–156. doi:10.1016/S0959-4752(02)00017-8.

Schnotz, W., & Kürschner, C. (2008). External and internal representations in the acquisition and use of knowledge: Visualization effects on mental model construction. *Instructional Science*, 36(3), 175–190. doi:10.1007/s11251-007-9029-2.

Tercedor, M., Alarcón-Navío, E., Prieto-Velasco, J. A., & López-Rodríguez, C. I. (2009). Images as part of technical translation courses: Implications and applications. *The Journal of Specialised Translation*, 11, 143–168. Retrieved from [http://www.jostrans.org/issue11/art\\_tercedoretal.pdf](http://www.jostrans.org/issue11/art_tercedoretal.pdf)

Tercedor-Sánchez, M. A., & Abadía-Molina, F. (2005). The Role of images in the translation of technical and scientific texts. *Meta: Translators' Journal*, 50(4). doi:10.7202/019857ar.

Tirkkonen-Condit, S., & Jääskäläinen, R. (Eds.) (2000). *Tapping and mapping the processes of translation and interpreting: Outlooks on empirical research*. Amsterdam: John Benjamins.

UTA käänöskomenttiohje. [Unpublished instructions for writing translation diaries for the translation students at the University of Tampere].

Wasylewsky, K., & Tapajna, N. (2001). The effects of positive and negative illustrations on text recall. *Ottawa Papers in Linguistics*, 29, 105–112. Retrieved from <http://artsites.uottawa.ca/clo-opl/doc/The-Effects-of-Positive-and-Negative-Illustrations-On-Text-Recall.pdf>

Youngs, S., & Serafini, F. (2013). Discussing picture books across perceptual, structural and ideological perspectives. *Journal of Language and Literacy Education*, 9(1), 189–200. Retrieved from <http://jolle.coe.uga.edu/wp-content/uploads/2013/05/Discussing-Picturebooks.pdf>

## **About the author**

Anne Ketola is a PhD candidate at the University of Tampere, Finland, and a member of the MULTI (Multimodality in Translation and Interpreting) research group, investigating the role of multimodality within translation studies. Her doctoral thesis examines the interaction of visual and verbal information during the translation of illustrated technical texts.

**Email.** anne.ketola@uta.fi

**URL.** [http://www.uta.fi/ltl/yhteystiedot/henkilokunta/ketola\\_anne.html](http://www.uta.fi/ltl/yhteystiedot/henkilokunta/ketola_anne.html)

### **Contact.**

University of Tampere  
Pinni B building, Room B5044  
Kanslerinrinne 1  
33100 Tampere  
Finland

Manuscript received May 21, 2015; revised June 6, 2015; accepted June 18, 2015.

Anne Ketola

## An Illustrated Technical Text in Translation

### Choice Network Analysis as a Tool for Depicting Word-Image Interaction

#### *Abstract*

The present study inquires into word-image interaction in the translation process of illustrated technical texts. The method employed for this purpose is Choice Network Analysis, which compares the translations of the same source text by multiple translators in order to empirically derive *the options*, the set of possible solutions, that are available when translating each verbal item. The data of the study consists of eight translations of an illustrated technical text, produced by a group of Master's level translations students. The study sets out to assess if the options offered by the multimodal source text are based entirely on verbal information or on a negotiation of meaning from two different modes. The analysis implies that visual information could modify verbal information; in the most extreme cases, visual information could cause verbal information being disregarded altogether. This indicates that images can reattribute the meaning of verbal items in translation.

## 1 Introduction

Illustrated texts consist of words and images and are hence multimodal: their message is created in the interplay of two separate modes. Even though translating illustrated texts is commonplace for translators, translation studies have until very recently only examined the verbal dimension of translation (O'Sullivan 2013: 2-3), insinuating, perhaps, that translators only process verbal information. However, a broad range of research has established that when reading an illustrated text, readers process both words and images, and form their interpretation of the multimodal text based on both verbal and visual information (see e.g. Connors 2013; Youngs/Serafini 2013; Schnotz/Kürschner 2008; Schnotz/Bannert 2003; Mayer 2002, 2005; Wasylensky/Tapajna 2001; Hegarty/Just 1993). As translators, too, start their work as readers, one might suggest that the same argument holds for a translator's interpretation of a multimodal source text combining words and images. However, the argument remains empirically unexplored.

The present study examines how translators process the combination of words and images – referred to in this study as *the verbal source text* and *the visual source text* – when translating illustrated technical texts. The definition of an illustrated technical text

adopted in the article builds on Byrne's description of technical texts: *illustrated technical texts* here refer to informative, instructive texts that explain how something works by both verbal and visual means (Byrne 2012: 26-28). Even though technical texts constitute a significant number of the texts being translated today and even though illustrations are an integral feature of technical documents (Byrne 2012: 6, 26, 54), research into the translation of illustrated technical texts has mainly focused on providing criteria for appropriate image selection in technical and scientific texts (Tercedor-Sánchez/Abadía-Molina 2005) and technically oriented multimodal terminological databases (Prieto Velasco 2009, 2012). Translation studies has yet to assess if and how images and image-word interaction are involved in the translator's interpretation of the source text and, consequently, translation solutions.

This article sets out to test one possible method of inquiring into the effect of word-image interaction on the translator's choices and the cognitive processing underlying the translation of illustrated texts, namely Choice Network Analysis (CNA), introduced by Campbell (2000a,b; see also Campbell/Hale 2002). CNA operates on the principle that comparing the translations of the same source text by multiple translators enables us to determine the different ways in which the same text may be interpreted. In CNA, different translation solutions are collected into a network-like flowchart in order to identify the similarities as well as the differences between them. The translation solutions that were employed by the particular group of translators are referred to as *options*. In the framework, the options are taken to represent the possible translation solutions that were available to the translators when translating a specific source text item (Campbell/Hale 2002: 18). This idea may be complemented by adding that it is obviously possible that the translators considered other possible solutions before settling on a particular option. One could therefore postulate that the options, the translation solutions that were actually employed by the translators, represent the translation solutions that the translators considered to be the most appropriate for the given context.

Applying CNA to a number of translations of the same illustrated text reveals the options that the multimodal source text offered for the translators, or the range of the possible ways in which the particular group of translators extracted information from the combination of words and images. This study is interested in examining if the multimodal source text – the verbal text combined with the illustrations – offers options that the verbal text on its own does not. In other words, the study aims to determine if the images can affect the way in which the verbal text is translated. The study is a part of a larger research project investigating illustrated technical texts. The following section provides an overview of the research project, outlining its background and objectives and summarizing its findings so far.

## 2 Overview of the Research Project

The study presented in this article is a part of a research project investigating how visual information is involved in the translation process of an illustrated technical text. The

research data of the project comprises two parts. The first part consists of eight translations of an illustrated technical text, analyzed in this article. The second part of the data consists of translation diaries – individual, reflective accounts on the problems encountered during the translation assignment, the strategies employed to solve these problems, and so on – analyzed in a previous article (Ketola 2015). The data was produced by a group of Master's level translations students at the University of Tampere, Finland, during a specialized technical translation course from English to Finnish. The students were given one week to complete the translation assignment. In order to guarantee an unaffected reaction towards the images, the students were not informed of the aim of the research prior to the translation assignment. The use of dictionaries and parallel texts was allowed and encouraged. After the students handed in their translations and translation diaries, a group interview was conducted, during which the aim of the research was made known.

The source text for the translation assignment presented the illustrated operating principles of two different types of wet magnetic separation devices, concurrent and counter-current wet magnetic drum separators, used in the mining industry for ore beneficiation. The source text was presented as a chapter of a mining engineering text book. It consisted of just over 500 words and two large colored images. In order to assess just how much the visual information may guide the translation process, the relationship between the words and the images of the source text had been modified in certain parts so that the information conveyed by the two modes is, in one way or another, *asymmetrical*. These modifications can be seen in figure 1.

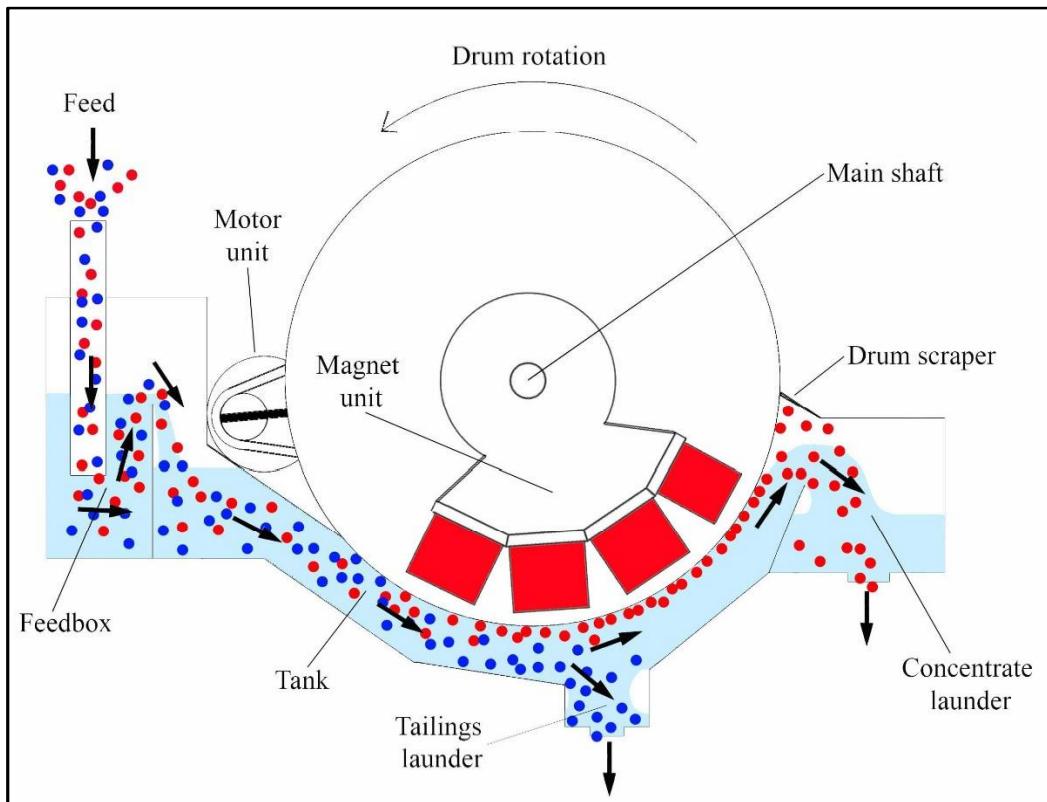


Fig. 1: The first image of the source text, representing the operating principle of a concurrent wet magnetic drum separator.

In two parts of the source text, the words and images of the source text were modified so that the information provided by the two is contradictory: the verbal text states that the drum of the device (the large circular part in the middle of the device) is submerged under water, while only the bottommost part of the drum is under water in the image. Further, the verbal text states that a part of the device, a *tailings launder*, is located in the *upper* part of the device while, according to the image, the part is located at the bottom of the device. Some of the parts of the machine were also reshaped in the image so that they no longer corresponded to their verbal definitions. An example of this is a part called *launder*: while the term typically refers to a trough or a narrow channel, the launders of the equipment are presented in the image as nearly square.

Further, in the second image of the source text (not displayed in the present article), some of the visual information was deleted from the image: the verbal text accurately described a particular part of the operating process (tailings or nonmagnetic particles exiting the separator) but the corresponding information could not be found in the image. These modifications were made in order to make it easier to distinguish which mode the translators based their translations on; had the visual and verbal modes expressed precisely corresponding information, it would have been considerably more difficult to

determine which mode the translator considered to be of more relevance during translation.

Ketola (2015) analyzed the translation diaries produced during the particular translation assignment. Even though the students were not instructed to inspect or comment on the images of the source text, issues related to the images, the asymmetry of information between words and images in particular, were widely discussed in the diaries. Only one of the translators did not discuss the images at all in the translation diary. Yet, in the group interview, this translator reported having studied the images before reading the verbal source text for the first time, but having disregarded the images when performing the actual translation. As for the rest of the group, the analysis of the translation diaries indicated that visual and verbal information challenging or contradicting each other had disturbed the translation process considerably. The analysis concluded that the students had repeatedly inspected the images during the translation assignment, which could be interpreted as visual information constituting a relevant part of the source text. The present article now sets out to determine which solutions they actually employed in their translations.

### 3 Choice Network Analysis

Choice Network Analysis (CNA) compares the translations of the same source text by multiple translators in order to construct models of the cognitive processing underlying translation. CNA is introduced both as an alternative and a complement to other research methods used for inferring these models, such as think-aloud protocols which, while being rich sources of information, are unable to access processing that is automatized (Campbell 2000a: 30) or, one might add, otherwise beyond conscious attention. CNA builds on the premise that while we cannot directly observe cognitive processing, we can infer a model of this processing from its results, meaning translations (Campbell 2000a: 33).

In CNA, a network-like flowchart, or a *choice network*, is constructed by comparing and classifying the different translation solutions produced by a number of translators for a source text segment. The size of the analyzed segment is determined by the aim of the research (Campbell 2000a: 38): it can be an individual word (e.g. Campbell/Hale 2002: 21-22), a string of several words (e.g. Campbell 2000a: 34) or even an entire phrase (e.g. Pavlović 2007: 164-166). Figure 2 represents an example choice network from the data (discussed in more detail later in the article) which displays the translations for the prepositional phrase *through an overflow*.

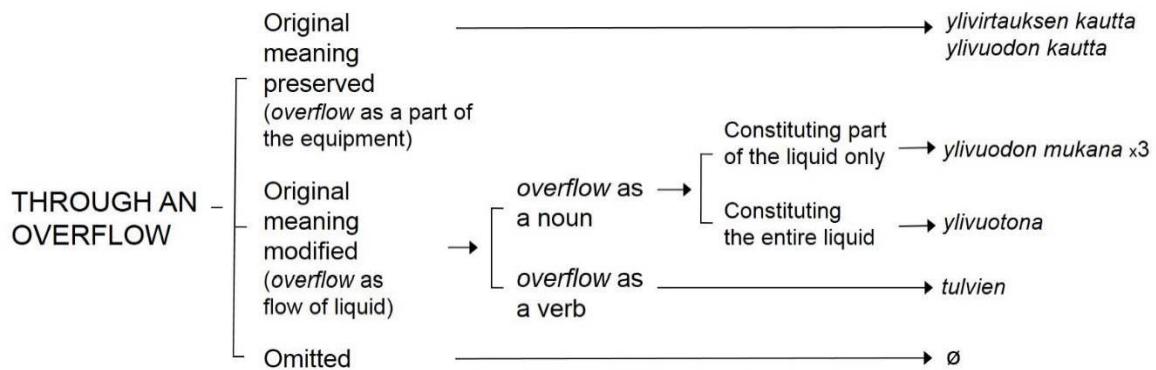


Fig. 2: Choice network of the prepositional phrase *through an overflow*.

The choice network represents all the solutions produced by the particular group of translators or *the options* that the source text offered for the group. The network can therefore be used to make general inferences about what went on in the translators' minds or, as Campbell puts it, "the processes that typically operate in particular type of subjects translating particular kind of texts under specific conditions" (Campbell 2000a: 31). These inferences can subsequently be used to construe more general principles about cognitive processing or to be used as hypotheses in the examination of other texts (Campbell 2000a: 32). An obvious limitation of CNA, as remarked by Pavlović, is that CNA on its own only displays the various possible results of the translators' decision-making process but not the reasons behind these solutions (Pavlović 2007: 178-179). However, this is compensated for in the present article – at least in part – by comparing the choice networks with the insights provided by the analysis of the translation diaries, in which the translators reflect upon the choices they made during translation.

Applying the method to the data of the present study enables two things: First of all, it allows to establish which options this particular multimodal source text offered for these particular translation students, or to establish "the potential pathways" (Campbell 2000a: 33) that the multimodal source text offered for the group. The close comparison of the different translation solutions invites us to reflect on the motives behind the solutions. The present article is interested in considering if the translation solutions seem to be based on purely verbal information or on a negotiation of meaning from two different modes, in other words, considering if the word-image interaction in the source text offered options not available from the verbal text alone. Secondly, on a more general level, applying the method to the data of the study enables us to construe more general principles about the cognitive processing taking place during multimodal text comprehension. The analysis reveals the range of the possible ways in which the particular group of translators interpreted the source text, in other words, extracted information from the combination of words and images. For this reason, the analysis may provide detailed empirical information about how the comprehension of a multimodal text unfolds – information which surely has applications even outside the immediate context of translation.

## 4 Data Analysis

This section of the article presents the analysis of the data. First, the section introduces a selection of five choice networks that are representative of the aims of the paper and discusses how they may be interpreted. The section then recapitulates the results of the analysis, compares them to the observations made during the analysis of the translation diaries and discusses what may be inferred about the cognitive processing underlying the translation of this particular multimodal text. The final part of the section evaluates the practicability of the method by discussing its advantages and disadvantages.

Campbell does not provide instructions on how the choice network analysis of a text segment should be performed in practice – the execution may obviously be designed to best suit the aims of the analysis. As mentioned above, choice network analyses in previous research have ranged from individual words to entire phrases. For the purposes of the present research – in which the individual translation solutions are meticulously compared – the verbal source text was divided into relatively small segments; in fact, all of the networks introduced as examples represent either prepositional phrases or compound nouns. In most of these networks, each segment has been further divided into smaller items: either individual words or chains of a few words. In practice, this division was decided on in a case-by-case basis, depending on what seemed logical when taking into consideration the structure of both languages involved. Each example first presents the phrase from which the analyzed segment has been extracted, with the analyzed segment in bold, followed by the choice network constructed of its translations.

### 4.1 Choice Network Analyses of the Data

**Example 1:** The magnetic particles are separated from the rest of the stream as they adhere to the drum surface **in the area of the magnet**.

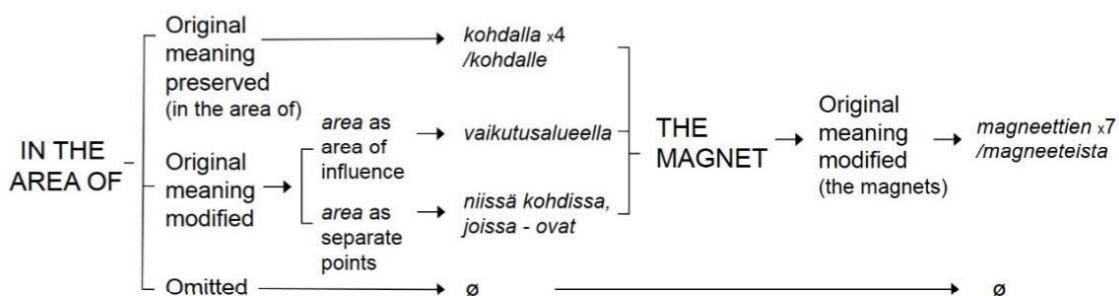


Fig. 3: Choice network of the translations of the prepositional phrase *in the area of the magnet*.

Figure 3 illustrates how the visual source text may affect the choice between singular and plural forms of nouns. The figure presents the choice network of the translations of the prepositional phrase *in the area of the magnet*. In other words, the network presents the options that the multimodal source text offered for the translation of the particular

prepositional phrase. In the choice network, the prepositional phrase has been divided into two items, *in the area of* and *the magnet*, which are analysed separately. The choice network of the prepositional phrase shows that one of the translators omitted the prepositional phrase altogether.

When looking at how the first item was translated, we can see that five translators employed a translation solution that could be described to preserve the original meaning of the item (*kohdalla* by four translators and *kohdalle* by one, differentiated only by the Finnish case suffixes [-lla, -lle] which correspond to prepositions). Two translators modified the meaning of the item: one translator formulated the item as *vaikutusalueella* ‘in the area of influence [of the magnets]’. This translation solution may be considered to explain the physical operating principle of the magnet slightly further and it therefore reflects that the translator comprehended this part of the operating process well. The other translator formulated the item as *niissä kohdissa, joissa [magneetit] ovat* ‘in the points where the [magnets] are’. The translation solution is a little misleading, since it does not convey the idea of a continuous magnetic field generated around the individual magnets. Comparing the translation solution with how this part of the process is presented in the image (see figure 1), one could claim that the solution reflects particularly well how the translator interpreted the verbal information by comparing it to the information provided by the image; an argument which may be supported by asking if the translator would have ended up with this solution had the source text not contained any images.

The right-hand side of the choice network displays the translation solutions for the latter item, *the magnet*. Of particular interest in this part of the network is the choice between singular and plural forms of nouns. The word is in the singular form in the verbal source text since it refers to the magnet unit of the separator as a whole. Yet, the visual source text clearly shows how the magnet unit consists of four individual magnets, colored with the same bright red as the dots representing the magnetic particles, hence emphasizing their connectedness. The choice network shows that all seven translators who preserved the prepositional phrase translated *magnet* in the plural form. It hence seems that the most obvious option the multimodal source text offered for the translators was in fact referring to the magnet in the plural form. The reason for this could well be that the visual source text so clearly depicts a series of four individual magnets. Again, it is quite unlikely that the translators would have ended up with this choice of words by interpreting verbal information only.

A similar example of how the visual source text may affect the choice between singular and plural forms of a noun could be seen in the choice network of a phrase describing how the ore particles are washed off the surface of a drum with *water sprays*. Yet, the corresponding part in the visual source text (not displayed in the present article) shows only one water spray showering water on the drum. The choice network of the phrase indicated that the multimodal source text offered three different options for the translation of *water sprays*: one translator omitted it, three translators translated it in the plural form (preserving the original meaning), and four translated it in the singular form.

**Example 2:** The weakly magnetic and non-magnetic particles are carried forward by the stream and eventually discharged from a tailings launder **in the upper part of the equipment.**

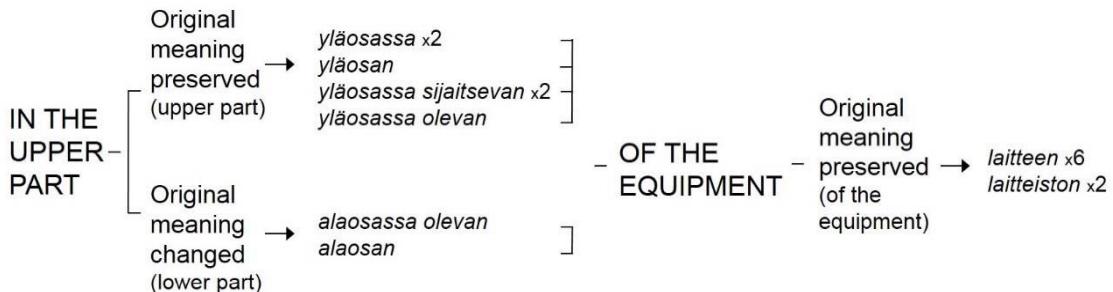


Fig. 4: Choice network of the translations of the prepositional phrase *in the upper part of the equipment*.

Figure 4 represents a part in the multimodal source text in which the relationship between the two modes was modified so that they provide contradictory information: according to the verbal source text, *the tailings launder* is located in the upper part of the equipment while, according to the visual source text (figure 1), the part is one of the bottommost parts of the device. As could, perhaps, be expected, the multimodal source text offered the translators two opposing options as to the location of the launder: the upper or the lower part of the device. The choice network shows that six translators decided to preserve the original meaning of the item, in other words, convey the information offered by the verbal source text (*yläosassa* ‘in the upper part’, *yläosan* ‘of the upper part’ or similar). Two translators radically restructured the original meaning of the item by conveying the information offered by the visual source text (*alaosassa olevan* ‘located in the lower part’, or *alaosan* ‘of the lower part’). This is quite an obvious example of the multimodal source text offering options that the verbal text on its own does not.

The fact that only two translators conveyed the location of the part as expressed by the visual source text is somewhat surprising taking into consideration that most translators (all seven who actually employed visual information during translation) discussed this contradiction of information in their translation diaries and most commonly reported it as a mistake in the verbal source text. The fact that they did not “correct this mistake” in their translations could well be due to insecurity: as several of them commented in their diaries, they felt like their background knowledge of the somewhat complex technical topic did not entitle them to change the verbal text. In other words, even though they identified the item as mistaken, they seemed to esteem the integrity of the verbal source text too much to change the item to what they in fact believed was correct.

The second example of contradiction between word and image had to do with the verb *submerge*: the verbal text expressed that the drum of the device is “submerged in a tank” – in other words, sunk below the surface of water – while, according to the image,

only the bottommost part of the drum is under water (see figure 1). The choice network of the particular phrase revealed that the multimodal source text offered basically two options for the verb *submerge*, one of which conformed to the verbal source text item somewhat literally (*upottaa* ‘submerge’, ‘sink’). The other, only employed by one of the translators, expressed that the drum is ‘placed’ in a tank (*sijoittaa*), which is roughly the same idea on a more general level. According to the analysis of the translation diaries, only two translators noticed the contradiction, one of whom decided not to change the wording. The other six translators affirmed in the post-translation group interview that they had not noticed the contradiction. This illustrates that the process of multimodal meaning construction is, in fact, quite indeterminable: the multimodal source text may offer differing options but these options are not always noticed by the translator.

**Example 3:** The non-magnetic particles flow in the opposite direction to the drum rotation and are discharged **through an overflow** into a tailings chute.

Figure 2, presented above in section 3, illustrates how the lack of visual information may affect the way in which the translator interprets the verbal source text. The example represents a part in the source text describing how the stream of the non-magnetic particles exits the separator after being separated from the stream of the magnetic particles; the two streams continue their journeys in opposite directions and exit the separator from separate outlets. While the verbal source text accurately describes the particular part of the operating process, the corresponding information cannot be found in the image (not displayed in the present article): the non-magnetic particles are not depicted in the image beyond the point where they separate from the magnetic ones. The *overflow* or the outlet through which the particles exit the separator can be seen in the image, but unlike in the first image of the source text (figure 1), the names of the parts of the equipment have not been labeled in the image corresponding to this part of the operating process. Moreover, the *tailings chute* (analyzed below in example 5) through which the particles continue their journey is not depicted in the image at all.

Figure 2 represents the choice network of the prepositional phrase *through an overflow*. The term *overflow* can refer to both the flowing over of a liquid (either as a noun or a verb) or, as in this particular text, an outlet for excess liquid. As the network displays, the source text offered the translators six different options for the translation of the prepositional phrase. Two of the translators produced a solution which conformed to the original meaning of the phrase (the uppermost solutions in the figure), in other words, convey the idea of an overflow as a part of the equipment. Both *ylivirtaus* and *yliuoto* (which take the forms of *ylivirtauksen* and *yliuodon* after being conjugated to the genitive case required by following postposition *kautta* ‘through’) are synonyms which refer to a passage in a machine through which excess liquid is let out. However, five translators employed a translation solution which describes *overflow* as flow of liquid, either as a noun or as a verb (e.g. ‘discharged as overflow’, ‘by overflowing’). While these translation solutions are not necessarily erroneous, by failing to mention the particular part of the equipment they do not describe the operation process as accurately as could be

expected of a mining engineering text book. Taking into account that the particles are not depicted as exiting through this particular part of the equipment and that the names of the parts of the equipment were not labeled in the particular image, one could suggest that due to the lack of visual information, the translators had trouble understanding what *overflow* referred to exactly in this particular text. As may be seen from the choice network, one translator omitted the prepositional phrase altogether. This, too, could be indicative of having had trouble understanding the phrase.

**Example 4:** The weakly magnetic and non-magnetic particles are carried forward by the stream and eventually discharged from **a tailings launder** in the upper part of the equipment.

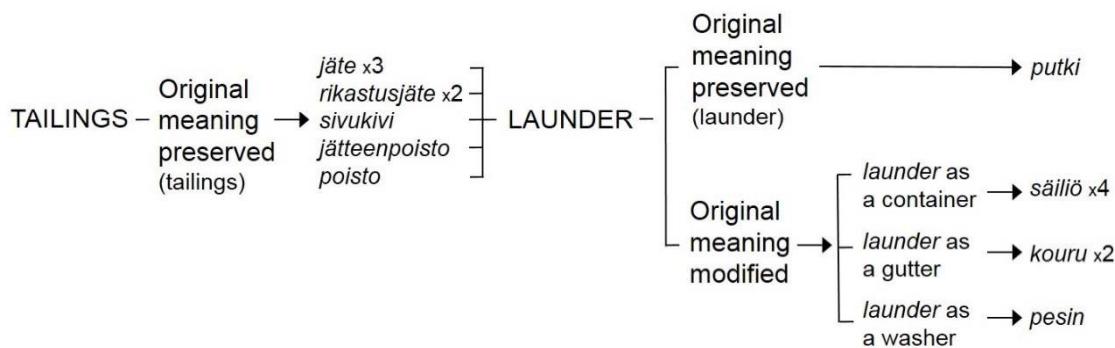


Fig. 5: Choice network of the translations of the term *tailings launder*.

The term *launder* is repeated various times in the source text because there are three separate launders in the separators introduced. In the subject area, the term typically refers to a long narrow channel or a trough used to convey liquids in a vertical direction. Yet, the shape of the launders was modified in the visual source text: they are depicted in the image as slightly wider than they are tall (see, for instance, the *concentrate launder* and *tailings launder* in figure 1). Figure 5 presents the choice network of the translations of term *tailings launder*. The translations for *tailings* are all acceptable near-synonyms and will not be analyzed here because they are not related to the research question of the article. As may be seen from the network, the multimodal source text offered four different options for the translation of *launder* (the exact same translation solutions were employed for the other launders mentioned in the source text as well). Only one of the options (*putki* ‘pipe’) clearly preserves the meaning of the verbal source text item; in other words, it describes a trough which could convey liquid in a vertical direction. Perhaps not surprisingly, this translation solution was employed by the translator who reported not having looked at the illustrations during translation.

The remaining seven translators employed a translation solution that did not entirely preserve the original meaning of the verbal source text item. Four of them translated the term as *säiliö* ‘container’. In the translation diaries, one of them explained the rationale behind the translation solution, stating that “container” was a simple choice of words

"since it is so generic it can be of any shape at all" (Ketola 2015: 28). Here the suitability of the translation solution is assessed in relation to the shape of the part in the visual source text. The problem with this translation solution is that a container is used to contain or hold something; the term does not convey the idea of channel through which liquid passes. One could hence suggest that for these four translators, it was more important to make the translation solution conform to the shape of the part in the image than to conform it to its function. Two of the translators translated the term as *kouru*, the closest English translation for which would be "gutter". *Kouru* commonly describes a groove or a furrow that is open from the top (Grönros et al. 2012). Since the part of the equipment in question conveys water in a *vertical* direction, this translation may be considered as unfit for the given context.

Lastly, one translator ended up translating the term as *pesin* 'washer', which is a rather odd choice of words for the given the context. It has most likely been selected straight from a dictionary, as *laundering* may also refer to the act of washing something. It is interesting to consider, however, if the image has supported this somewhat unsuitable translation solution. The liquid (finely ground ore mixed with water) moving inside the magnetic separator has been depicted in the image as light blue while, in reality, it is a dirty shade of brown. It would be interesting to know whether the translator would have ended up with the particular solution even if the image had depicted the liquid in a colour that is not as easily associated with the state of being clean. To sum up, modifying the shape of the part in the image so that it no longer resembled a trough resulted in that none of the translators who actually compared the term with how it was depicted in the image while translating employed a translation solution literally describing a trough.

**Example 5:** The non-magnetic particles flow in the opposite direction to the drum rotation and are discharged through an overflow into a **tailings chute**.

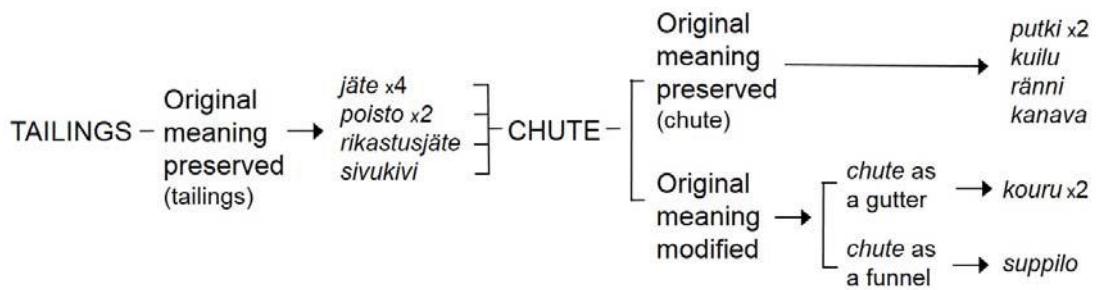


Fig. 6: Choice network of the translations of the term *tailings chute*.

Figure 6 represents the choice network of the translations of the term *tailings chute*. As mentioned in the discussion of example 3, analyzing a preceding part of the same phrase, the *tailings chute* is not depicted in the corresponding image at all. Earlier in the article it was suggested that the translators might not have understood the particular phrase correctly due to the lack of visual information.

The choice networks for the terms *tailings launder* in figure 5 and *tailings chute* in figure 6 offer an interesting comparison of how the visual source text may affect terminological choices during translation. The two parts of the equipment are identical in function and usually highly similar in appearance: they are both vertical troughs which convey liquid containing the tailings (non-magnetic ore particles). In fact, employing the exact same term to describe them both would be recommendable for the sake of terminological consistency. However, different terms were employed in this source text in order to elucidate whether the differences in the corresponding visual information would cause them to be translated differently.

As may be seen in figures 5 and 6, the options of *putki* and *kouru* were the only two that were employed for both terms. Unlike in the previous example, nobody translated *chute* as ‘container’, which further affirms the idea that it was, in fact, the visual source text that prompted this option. Compared to the translation solutions presented in the previous example, the translators provided four novel solutions: *kuilu*, *ränni*, *kanava*, and *suppilo*. The first three of these can be considered to more or less preserve the meaning of the original, even though *kuilu* can also refer to a ‘shaft’ and *kanava* to a ‘channel’. *Suppilo* ‘funnel’, employed by one translator, refers to a tube or pipe that is wide at the top and narrow at the bottom. Perhaps not surprisingly, this option was not employed in the previous example where it would have contradicted the visual source text. In other words, *chute* offered the translators options that *launder* did not. One could suggest that not limiting the meaning of the verbal text with visual information opened it up for new interpretations. Conversely, this would mean that visual information can restrict the translators’ interpretation of verbal information, or specify the meaning of verbal items in translation.

#### 4.2 Discussion of the Results

The article will now consider what the choice networks of the data may disclose about the cognitive processing underlying the translation of this particular multimodal text, and how the analysis of the translation diaries (Ketola 2015) may further elucidate these observations. The networks introduced above represent the range of options that a particular segment of the multimodal source text offered for the group of translators. What these choice networks have in common is that some of the options displayed in each seem to be affected by the information conveyed by the visual source text – the combination of the visual and verbal information offered options that the verbal source text on its own did not.

The comparison of the translation solutions for the word *magnet* (figure 3) reveals that the multimodal source text seemed to offer only one obvious option which was employed by all of the translators who actually included the particular word in their translations. Curiously enough, the choice between singular and plural noun form corresponded to the information provided by the visual source text. It would therefore seem that the visual source text affected the way in which this item was translated. One could hence suggest that the cognitive processing of the item involved interpreting both

verbal and visual information. This part of the source text was not discussed in the translation diaries, perhaps because the changes made were so minor that the translators did not feel the need to justify them, or perhaps because the translators were not consciously aware of these changes.

The choice network of the prepositional phrase *in the upper part of the equipment*, figure 4, showed that some translators disregarded the information conveyed by the verbal source text altogether, which is perhaps the most obvious example of the visual source text guiding the translation solution found in the data of the study. What can be inferred about the cognitive processing behind this item? Based on the choice network alone, one may conclude that at least the two translators who decided to reproduce the information provided by the image processed both verbal and visual information – to the degree that they noticed the contradiction of information. In addition, based on the analysis of the translation diaries, one may conclude that the same was true for five others as well, even though they decided to preserve the information as expressed by the verbal source text. All in all, it could be suggested that the cognitive processing of the particular item generally included a careful, intentional negotiation between the information provided by the two modes. However, this is not to say that the fact that the visual mode provides information that is contradictory to that provided by the verbal will always result in the translator having to choose between one or the other: as illustrated by the example of the verb *submerge*, the translator does not always inspect both modes thoroughly enough to acknowledge a contradiction of information.

Based on the choice network alone, it is quite challenging to interpret the translations of the prepositional phrase *through an overflow* (figure 2). Yet, the analysis of the translation diaries considerably elucidates the translators' thought processes. Most of them explicitly reported not having understood this particular phrase. The study (Ketola 2015: 27) suggested that instances such as these reflect an effort to interpret the multimodal source text as a whole: the translator maps visual and verbal information onto each other, or confirms the verbally acquired information from the image. If this cannot be done (as in the present example), the translator is left with an impression of not having understood the verbal text at all. This information is valuable for the interpretation of the choice network. Only two of the options displayed by the choice network preserve the original meaning of the item (*overflow* as a part of the equipment). The rest of the options – which either conveyed a somewhat mistaken idea of the item or omitted it – reflect the fact that the translator had trouble understanding the meaning of the item. One can suggest that the cognitive processing of the item included negotiating the combination of verbal and visual information and that the lack of visual information complicated this processing considerably.

Another example of how visual information – or the lack of it – may affect translation may be seen in figures 5 and 6 representing the translations of the terms *launder* and *chute*. As stated above, the terms represent two functionally corresponding parts of the equipment, designated in the verbal source text by two near-synonyms. What sets them apart is how they are depicted in the visual source text. The translation solutions

produced for *launder* seemed to conform to the way in which the part was depicted visually more than to the way in which it was described verbally. In turn, the somewhat inaccurate translation solutions produced for *chute* reflected that the translators had trouble understanding what the term refers to exactly. Again, one could claim that the cognitive processing of the multimodal source text has included thoroughly inspecting both modes: when visual information was available, it affected the way in which the verbal text was interpreted. When visual information was not available, the inability to interpret the modes in relation to each other complicated the translators' efforts at arriving at a translation solution.

All in all, the analysis of the data implies that the image may refocus the interpretation of the verbal text. Interpreting verbal information in relation to images may result in a verbal source text item being attributed meanings it would otherwise not have, or even disregarding verbal information altogether and replacing it with information conveyed by the visual. The translator may attempt to interpret words in relation to images to the extent that the lack of visual information and, consequently, the inability to interpret the modes in relation to each other, disturbs the translators' source text comprehension. However, as the chosen method of analysis readily displays, the same combination of verbal and visual information may be interpreted in various ways. The fact that an image has the potential to specify the meaning of a verbal source text item does not automatically mean that this specification of meaning always takes place. In the same fashion, the fact that the verbal and the visual source texts provide contradictory information does not necessarily mean that the translator purposely chooses between one and the other.

#### 4.3 Evaluation of the Method

As illustrated by the above discussion, CNA can be a valuable component when examining the range of translation solutions that a source text may offer for a group of translators. However, it is reasonable to ask how much we may really infer of a translator's cognitive processing simply by examining the results of this processing. As posited earlier in the article, CNA is not able to access the reasons behind the translation solutions – which, admittedly, is an inherent limitation of all product-based studies investigating the cognitive processing during translation. This limitation also repeatedly became evident in the present analysis: for instance, in the analysis of the part of the source text in which the two modes provided contradictory information, CNA only revealed that the majority of the translators conveyed the information expressed by the verbal text. Without the analysis of the translation diaries we would not know that the majority of the group actually identified the verbal element as mistaken, but nonetheless insisted on conveying this information to their translations. Without the analysis of the translation diaries, in other words, we would not know that both of the modes were attentively scrutinized and that the choice between either conveying the information provided by the verbal or the visual mode was negotiated with conscious attention. One

may therefore conclude that it is highly recommendable to triangulate CNA with a complementary method of acquiring into how the translation process unfolds.

Despite its limitations, the method yields a useful tool for the empirical investigation of translation. Laying out the range of translation solutions that were produced for the same source text segment is an excellent way of illustrating in how many different ways the same segment may be interpreted and translated. Further, it encourages the close comparison of the translation solutions and hence invites the analyst to contemplate on the reasons behind the differences in the solutions. In practice, the detailed construction of the networks is somewhat time-consuming and the author would therefore suggest that the method is best suited for relatively small samples of data, such as the data of the present study. Campbell's own applications of the method, too, usually employ a sample size of nine participants (Campbell 2000a: 34, 2000b: 218; Campbell/Hale 2002: 23). Limiting the number of participants is obviously another limitation of CNA, since the observations made during the analysis can, by no means, be claimed to be exhaustive. A larger sample of participants would most likely reveal yet more ways of interpreting the source text.

## 5 Conclusions

The article set out to assess if the images of an illustrated technical text can affect the way in which the verbal text is translated. Based on the analysis of the data, we may postulate that the image is capable of reattributing the meaning of a verbal element. For instance, visual information can specify verbal information: some of the translation solutions in the data could reasonably be claimed to be influenced by the way the corresponding information was represented in the visual source text. In the most extreme cases, conflicting visual information may cause verbal information to be disregarded altogether. All in all, the meaning of the multimodal source text item often seems to be negotiated from the combination of both modes. If visual information is not available, the attempt to negotiate meaning from the multimodal ensemble may even disturb the translators' source text comprehension.

Yet, making claims about how the interpretation of verbal information is affected by the visual means that we are looking at the subject from a verbally informed perspective; the interpretation of the image is obviously also affected by verbal information. From a more holistic point of view, we may postulate that interpreting the combination word and image may result in a novel interpretation of the whole: verbal and visual information may redetermine each other and hence, together, they may codetermine the meaning of the whole.

It must be emphasized that the combination of verbal and visual information can be interpreted in various different ways: the meanings the two modes produce together are neither predetermined nor clear-cut. The combination of word and image cannot be equated with a set meaning they automatically produce when presented together.

Moreover, multimodal integration of meaning does not always take place even if various modes are simultaneously presented for the reader (translator): the reader does not necessarily acknowledge both modes with equal attention at all times.

The discussion presented in the article hopes to emphasize the importance of examining the translations of illustrated texts in relation to both of the modes constituting the source text. On a more general level, the discussion hopes to further encourage translation studies towards examining how modes other than the verbal are involved in the translation process. As Campbell remarks, the observations made by CNA can subsequently be used as hypothesis in the examination of other texts (2000a: 32). Recommended directions for future research could include comparing choice networks to data collected using methods such as eye-tracking and keystroke logging. Eye-tracking methodologies could be used to record where the visual attention of the translator lies in each moment and hence determine how often the visual source text is actually inspected; coupled with keystroke logging this information could establish when exactly these inspections took place during the translation process.

**trans-kom**

**ISSN 1867-4844**

**trans-kom** ist eine wissenschaftliche Zeitschrift für Translation und Fachkommunikation.

**trans-kom** veröffentlicht Forschungsergebnisse und wissenschaftliche Diskussionsbeiträge zu Themen des Übersetzens und Dolmetschens, der Fachkommunikation, der Technikkommunikation, der Fachsprachen, der Terminologie und verwandter Gebiete.

Beiträge können in deutscher, englischer, französischer oder spanischer Sprache eingereicht werden. Sie müssen nach den Publikationsrichtlinien der Zeitschrift gestaltet sein. Diese Richtlinien können von der **trans-kom**-Website heruntergeladen werden. Alle Beiträge werden vor der Veröffentlichung anonym begutachtet.

**trans-kom** wird ausschließlich im Internet publiziert: <http://www.trans-kom.eu>

Redaktion

Leona Van Vaerenbergh  
University of Antwerp  
Arts and Philosophy  
Applied Linguistics / Translation and Interpreting  
Schilderstraat 41  
B-2000 Antwerpen  
Belgien  
[Leona.VanVaerenbergh@uantwerpen.be](mailto:Leona.VanVaerenbergh@uantwerpen.be)

Klaus Schubert  
Universität Hildesheim  
Institut für Übersetzungswissenschaft  
und Fachkommunikation  
Universitätsplatz 1  
D-31141 Hildesheim  
Deutschland  
[klaus.schubert@uni-hildesheim.de](mailto:klaus.schubert@uni-hildesheim.de)

## References

- Byrne, Jody (2012): *Scientific and Technical Translation Explained*. Manchester: St. Jerome
- Campbell, Stuart (2000a): "Choice Network Analysis in Translation Research." Maeve Olohan (ed.): *Intercultural Faultlines. Research Models in Translation Studies I*. Manchester: St. Jerome, 29-56
- Campbell, Stuart (2000b): "Critical Structures in the Evaluation of Translations from Arabic into English as a Second Language." *The Translator* 6 [2]: 221-229
- Campbell, Stuart; Sandra Hale (2002): "The Interaction Between Text Difficulty and Translation Accuracy." *Babel* 48 [1]: 14-33
- Connors, Sean P. (2013): "Weaving Multimodal Meaning in a Graphic Novel Reading Group." *Visual Communication* 12 [1]: 27-53
- Grönros, Eija-Riitta et al. (eds) (2012): "Kouru." *Kielitoimiston sanakirja*. Helsinki: Kotimaisten kielten keskus – <http://www.kielitoimistonsanakirja.fi/netmot.exe?motportal=80> (06.05.2016)
- Hegarty, Mary; Marcel A. Just (1993): "Constructing Mental Models of Machines from Text and Diagrams." *Journal of Memory and Language* 32 [6]: 717-742
- Ketola, Anne (2015): "Translation Diaries of an Illustrated Technical Text. Translation Students' Conceptions of Word-Image Interaction." *Connexions – International Professional Communication Journal* 3 [2]: 13-40 – <https://connexionsj.files.wordpress.com/2016/02/ketola.pdf> (06.06.2016)
- Mayer, Richard E. (2002): "Using Illustrations to Promote Constructivist Learning from Science Text." José Otero, José A. Léon, Arthur C. Graesser (eds): *The Psychology of Science Text Comprehension*. Mahwah: Erlbaum, 333-356
- Mayer, Richard E. (2005): "Cognitive Theory of Multimedia Learning." Richard E. Mayer (ed.): *Cambridge Handbook of Multimedia Learning*. New York: Cambridge University Press, 31-48
- O'Sullivan, Carol (2013): "Introduction: Multimodality as Challenge and Resource for Translation." *Journal of Specialized Translation* 20: 2-14 – [http://www.jostrans.org/issue20/art\\_osullivan.pdf](http://www.jostrans.org/issue20/art_osullivan.pdf) (06.05.2016)
- Pavlović, Nataša (2007): *Directionality in Collaborative Translation Processes. A Study of Novice Translators*. PhD Dissertation. Tarragona: Universitat Rovira i Virgili – <http://www.tesisenxarxa.net/TDX-1210107-172129> (06.05.2016)
- Prieto Velasco, Juan Antonio (2009): *Traducción e imagen: la información visual en textos especializados*. Granada: Tragacanto
- Prieto Velasco, Juan Antonio (2012): "A Corpus-based Approach to the Multimodal Analysis of Specialized Knowledge." *Language Resources and Evaluation* 47 [2]: 399-423
- Schnotz, Wolfgang; Maria Bannert (2003): "Construction and Interference in Learning from Multiple Representation." *Learning and Instruction* 13 [2]: 141-156
- Schnotz, Wolfgang; Christian Kürschner (2008): "External and Internal Representations in the Acquisition and Use of Knowledge: Visualization Effects on Mental Model Construction." *Instructional Science* 36 [3]: 175-190
- Tercedor-Sánchez, María Isabel; Francisco Abadía-Molina (2005): "The Role of Images in the Translation of Technical and Scientific Texts." *Meta: Translators' Journal* 50 [4] – <https://www.erudit.org/revue/meta/2005/v50/n4/019857ar.pdf> (06.05.2016)
- Wasylensky, Krista; Nicole Tapajna (2001): "The Effects of Positive and Negative Illustrations on Text Recall." *Ottawa Papers in Linguistics* 29: 105-112
- Youngs, Suzette; Frank Serafini (2013): "Discussing Picturebooks Across Perceptual, Structural and Ideological Perspectives." *Journal of Language and Literacy Education* 9 [1]: 189-200 – <http://jolle.coe.uga.edu/wp-content/uploads/2013/05/Discussing-Picturebooks.pdf> (06.05.2016)

*Author*

Anne Ketola is a PhD candidate at the University of Tampere, Finland, and a member of the MULTI (Multimodality in Translation and Interpreting) research group, investigating the role of multimodality within translation studies. The present paper is a part of her doctoral thesis in which she examines the interaction of visual and verbal information during the translation of illustrated technical texts. Her other research projects investigate word-image interaction during the translation of other illustrated text types, such as picturebooks.

E-mail: [anne.ketola@uta.fi](mailto:anne.ketola@uta.fi)

Website: [http://www.uta.fi/ltl/yhteystiedot/henkilokunta/ketola\\_anne.html](http://www.uta.fi/ltl/yhteystiedot/henkilokunta/ketola_anne.html)

# Using Translation Research to Model Word–Image Interaction

Anne Ketola

University of Tampere, FINLAND

## ABSTRACT

This paper sets out to examine the potential of translation research as a means to model multimodal meaning construction in an empirical manner. The paper introduces two classifications of word–image relationships put forward in previous multimodality research (Marsh and White 2003; Martinec and Salway 2005). These are compared to empirical analyses from a research project examining the translation of an illustrated technical text. The paper suggests that the reviewed classifications do not cover all of the possible ways in which words and images can co-construct meaning: the translators interpreted the combination of words and images in various ways, some of which are clearly not represented in the classifications. The paper concludes that word–image interaction is more complex than the classifications suggest and that relying on ready-made classifications in our analyses of word–image interaction may limit the way we perceive the manifold nature of multimodal meaning construction.

**KEYWORDS:** Word–image interaction, multimodality, illustrated text, word–image taxonomies, technical illustrations

## 1. Introduction

In the last decades, a variety of disciplines, such as research in children’s literature, education, journalism and information design, have theorized *word–image interaction*, the way in which words and images in illustrated texts may interact and the relationships that may hold between the two modes. Yet, as John A. Bateman (2014:39–40) remarks, the observations made about the way words and images co-construct meaning are often simply theoretical assumptions of ‘what appears “obvious” or self-evident for the analyst’: they are based solely on a detailed analysis of the multimodal text by the analyst(s), and are generally empirically unsubstantiated.

Anne Ketola, *Using Translation Research to Model Word–Image Interaction*, 82 – 104.

In this paper, I propose that research into translation may offer a way to exemplify empirically how the combination of words and images is interpreted. The paper introduces empirical data from a research project in Translation Studies, examining how a group of translation students constructed meaning from the combination of words and images while translating an illustrated technical text under controlled conditions. This data includes the translations produced by the students (analyzed in Ketola 2016a) as well as the translation diaries the students wrote during the assignment (analyzed in Ketola 2015). This empirical research is compared to the way in which word–image interaction has been theorized in meta-research derived from other disciplines in order to reflect on whether these theories are supported by translation-specific empirical data. The aim of the paper is two-fold: it sets out to advance the conceptualization of multimodal meaning construction within Translation Studies as well as to examine the interdisciplinary promise translational enquiry offers for all research into multimodality.

In this paper, I propose that we may examine translation of multimodal texts to make inferences about the way in which these texts were interpreted; after all, translation presupposes interpretation. As emphasized by, for instance, Muñoz Martin (2010:175-6), it is not texts but individual interpretations that are transferred from one language to another in translation. When we examine a translation, we are, in fact, examining the interpretation of a particular text by a particular translator. If we examine the translation of a multimodal text, we are examining how a particular translator constructed meaning from the interaction of the modes. Studying the process of translation may therefore offer valuable insights into how we construct meaning from different types of texts. As Kussmaul aptly describes, we may examine translation

to show what goes on during normal language processing. Translations could then be regarded as a kind of ‘free production’ of a text we have read – albeit in different language. Translations would then be a kind of ‘reproduction experiment’ and could be used to investigate general language processing in the human mind (2000:69).

Analyzing the translations of an illustrated text by multiple translators could be viewed as ‘a reproduction experiment’ which allows us to compare the different ways in which meaning was constructed from the combination of the modes.

In this paper, I critically review two prominent meta-studies which present a classification of the possible ways in which words and images can co-construct meaning, and compare these to observations drawn from translation analyses. The first reviewed classification is an example of word–image relationship taxonomies (Marsh and White 2003), and the second is an example of accounting for word–image relationships by modelling them on the structure of the grammar of verbal language (Martinec and Salway 2005).

With the help of the empirical data, I propose that multimodal meaning construction is in fact more complex than the reviewed classifications suggest. This proposition includes two central arguments. In the first place, I argue that there are more possible word–image relationships than these classifications propose. The empirical examples of my study display that the combination of words and images can be interpreted in a variety of ways, some of which do not correspond to the possible word–image relationships described in the reviewed classifications. I therefore argue that the classifications do not represent all of the possible relationships that may hold between words and images. I also emphasize that conceptualizing word–image interaction by modelling it on simplified classifications may limit the way we perceive the complexity of this interaction.

In the second place, I argue that we cannot create classifications of word–image interaction that would allow us to label this interaction in a particular word–image pair in a conclusive way (for instance, define the relationship as *elaborating* or *enhancing*). In his evaluation of word–image relationship classifications, Bateman (2014:190) notes that analyses of word–image interaction can lead to situations where ‘different analysts come up with different descriptions of the same objects of analysis.’ He continues by arguing that ‘such variability is *not* due to multimodal artifacts exhibiting a rich array of meanings to be drawn out’ (2014:190 [emphasis in the original]), but instead are due to the classification system not being developed enough. In this paper, I argue that the opposite holds; that multimodal objects may indeed exhibit various meanings and hence be interpreted in varying ways.

The idea that the same combination of words and images can be interpreted in different ways is, by no means, a novel one. The notion is introduced as early as 1982 in Spillner’s seminal study on how a group of people interpreted an illustrated advertisement. Unfortunately, this diversity of interpretation is not always acknowledged in theoretical accounts of word–image interaction. Similar concerns have been raised by Holsanova (2012:252) who points out that

studies in semiotics and visual communication have ignored inter-observer differences when studying word–image interaction.

This paper is structured as follows. I start by critically introducing the two classifications and discuss why and how they might be limited in their applicability. I then introduce insights from my translation analyses regarding the way in which the group of translators interpreted the combination of words and images. I conclude by outlining how these insights may inform our understanding of the ways in which the two modes can co-construct meaning, as well as reflecting on the importance of further widening the scope of Translation Studies to a more multimodally-informed direction.

## 2. Modelling Word–Image Interaction

### 2.1. *Premises in Previous Research*

A widely adopted way of modelling word–image interaction is to propose a set of relationships that may hold between the verbal and the visual modes; ever since the work of Barthes (1964), there have been numerous attempts to arrive at an orderly classification of word–image relationships. I will now briefly introduce two such studies which represent distinct approaches. The first one is an example of word–image relationship taxonomies (Marsh and White 2003), meaning empirically deduced, classified lists of the possible relationships, and the other is an example of accounting for word–image relationships by modelling them on the grammar of verbal language (Martinec and Salway 2005). These examples have been chosen because both could be argued to be more comprehensive than other similar classifications. As discussed below, Marsh and White’s classification combines 24 taxonomies created in previous research. Martinec and Salway’s classification, too, has been complimented for being more exhaustive than others (Bateman 2014:197).

The theoretical starting points of these two classifications are substantially different. Judging by the lack of other justifications, taxonomies tend to be based on the analysts’ own empirical observations, whereas Martinec and Salway’s classification is constructed by mirroring word–image interaction on a ready-made set of possible relationships extracted from the study of verbal language. What these classifications have in common is that they both claim to be complete and to cover all illustrated text types. Martinec and Salway describe their

classification as ‘a generalized system of image–text relations. Our system aims to account, in a principled manner and in some detail, for *all the image–text relations in both new and old media*’ (2005:343 [emphasis added]). Marsh and White affirm that their taxonomy is applicable ‘to all subject areas and document types’, and argue for its completeness by emphasizing the elaborateness of their classification development process (2003:647–53). This paper sets out to question these notions, and suggests that words and images can also co-construct meaning in ways that are not represented in these classifications.

## 2.2. Modelling Word–Image Interaction as a Taxonomic Exercise

Numerous scholars in various disciplines have set out to create taxonomies of possible relationships that may hold between verbal and visual information. A thorough cross-section of such studies is offered by Marsh and White (2003), who compare and combine 24 taxonomies developed in various fields of research (children’s literature, education, journalism and information design, among others), including the work of prominent scholars such as Kress and van Leeuwen (1998), Levin and Mayer (1993) and Nikolajeva and Scott (2000). Marsh and White review the categories suggested by others and introduce a master taxonomy comprising 49 possible relationships between words and images, making it possibly the most comprehensive taxonomy proposed so far. The taxonomy suggests that words and images may, for instance, *complement* each other (one mode helps the other to convey a message), *contrast* each other (words and images oppose each other for rhetorical or narrative purposes) or *develop* each other (elaborate, specify or amplify each other’s meanings) (Marsh and White 2003:671–2).

Marsh and White’s (2003:654–60) example analyses reveal that the authors do not suggest these relationships as exclusive of each other. Instead, they assert that more than one relationship can hold in a particular word–image pair. On the one hand, I believe this should be considered as an advantage of the approach. It is, in principle, easy enough to imagine a situation in which an image could both help convey a verbal message as well as to elaborate it – hence both *complement* and *develop* the verbal mode. On the other hand, this flexibility of the categories also brings about questions of ambiguity: in a way, it emphasizes that the proposed categories might overlap to certain extent. One may also ask if the proposed categories are internally coherent. To give an example, if we conclude that a *developing* relationship holds between visual and verbal information, do the visual and the verbal information *specify* or *amplify* each other’s meanings (see description of the category above)?

Anne Ketola, *Using Translation Research to Model Word–Image Interaction*, 82 – 104.

Surely specification of meaning and amplification of meaning do not refer to the same thing. Do some readers interpret the relationship as amplification and others as specification? I propose that this ambiguity reflects how challenging it is to label a relationship in a way we could all agree on.

One can also reflect on the proposed applications of the classification. Marsh and White conclude that the classification can be used not only as a tool to design and create documents combining verbal and visual information, but also as a tool to ‘predict the effects of combinations [of words and images] once documents have been published’. *Predict* is a risky choice of words. One might interpret this as suggesting that a careful analysis of a word–image pair in a particular document would enable us to predetermine the way in which the interaction of the modes will be perceived by the readers of the document. Our analysis might conclude that a certain type of relationship (or relationships) exist between a word–image pair, but can we be sure that all readers interpret the relationship in the same way? If, for instance, an image helps one reader to assimilate a verbal message, will it inevitably help others as well? As Bateman affirms in his critique of taxonomies, no matter how detailed and organized our taxonomy lists are, suggesting such relationships between words and images must always be considered as ‘a *hypothesis* concerning interpretation’ (2014:48 [emphasis in the original]). Unless we set out to empirically examine how a proposed relationship is interpreted by different readers, we should be careful in making predictions about its nature.

### 2.3. Modelling Word–Image Interaction on Grammatical Categories

Martinec and Salway (2005) offer an example of classifying word–image relationships by modelling them on the structure of the grammar of verbal language. The classification, like various other classifications based on grammar, is presented in the form of a network of possible connections (see also e.g. Kong 2006; Matthiessen 2007). Martinec and Salway set out to develop what they call a generalized classification system of word–image relationships, based on reworking the grammatical concept of logico–semantic relations, originally introduced for verbal language by M.A.K. Halliday (1985). In addition to evaluating word–image relationships by modelling them on logico–semantic relations of verbal language, Martinec and Salway propose that word–image relationships are also realized on the level of status; words and images can either have an equal or unequal status. They thus describe the former status as a circumstance in which a whole image is related to a whole text, and the

latter as a circumstance in which one mode modifies the other (Martinec and Salway 2005: 345)<sup>1</sup>.

Logico-semantic relations refer then to the way clauses (and subsequently, words and images) can be thought to combine into more complex entities. Halliday's functional grammar distinguishes two main types of clause-combining relationships: the category of *projection* (which refers to the relationship between the event of talking and what is being said; in the context of word–image interaction, such relationship could be found in speech bubbles in cartoons) and *expansion*, which posits three different ways of 'adding' information. The types of expansion, as employed to the analysis of word–image relationships by Martinec and Salway (2005:351-54), are the following: *elaboration* (one mode adds information to the other, making its meaning more specific), *extension* ((one mode adds new, related information to the other; the combination goes beyond what is expressed in either mode alone) ) and *enhancement* (one mode adds qualifying information – related to time, place, manner, purpose, and so on – to the other).

Even though Bateman, quoted above, criticized taxonomies of word–image relationships for offering nothing but *hypotheses* concerning interpretation, he praises Martinec and Salway's classification. He suggests, for instance, that it may offer 'greater empirical adequacy and coverage' (2014:191) than other accounts of word–image relationships. Yet, one might ask if a grammatically-based network of possible connections between words and images is able to produce something beyond simple hypotheses. Why would a classification network, in its ability to predict the reader's interpretation, be that much different from a classification list? The improved empirical adequacy mentioned by Bateman most likely refers to inter-observer reliability. Yet, the validity of Martinec and Salway's analyses has been questioned, for instance, by Unsworth and Cléirigh (2009:153) who propose alternative interpretations for their analyses of examples.

My aim is, by no means, to claim that the relationships Martinec and Salway suggest for words and images cannot hold. On the contrary, as I also assert in the discussion of my examples below, some of the relationships they propose do seem to plausibly describe word–

---

<sup>1</sup> This dimension of Martinec and Salway's proposal is not discussed in this article because of space constraints, but one could subject it to similar criticism by asking if no other type of statuses can exist and by considering how inter-observer differences would affect defining the status.

image interaction under certain circumstances. However, a question that could be asked of the classification (as well as of Marsh and White's classification introduced above) is whether or not this classification provides a complete set of possible word–image relationships. Martinec and Salway's classification includes types of word–image relationships that correspond to the logico–semantic relations of verbal language. Relationships that do *not* correspond to these linguistic structures would thus fall out of the classification. The elegance of Martinec and Salway's classification might obscure the fact that other types of relationships could also exist between words and images. There is no a priori reason why word–image relationships should adhere to the relationships that may hold between linguistic clauses.

#### *2.4. Evaluation of the Classifications*

Martinec and Salway (2005:343) mention basing their research on electronic encyclopedias for children, textbooks, online news, printed ads and online gallery sites; Marsh and White (2003:647) report having analyzed educational material for children, online news and retail business pages. In other words, the data used in the studies have been somewhat similar. Moreover, one could claim that, as a general rule, the verbal and the visual messages in these sets of data have been meticulously articulated to work together: the modes have been orchestrated so that both offer largely compatible perspectives into their common referent – the object, event, or idea presented by multimodal means. What I mean by this is that the selection or the production of images, for instance, for a children's encyclopedia page is likely to have been the result of careful planning and a meticulous comparison of the verbal and visual messages involved. The same holds for a large portion of the illustrated material that is usually selected as data in research theorizing about word–image interaction: advertisements (e.g. Liu and O'Halloran 2009), picturebooks (e.g. Moya Guijarro 2014), comics and graphic novels (e.g. Connors 2013), and so on. If these types of texts contain instances of contradiction of information – asymmetry between the information conveyed by the verbal and visual modes – it usually serves a carefully calculated narrative function. In other words, it can convey humor and irony, or simply call for the reader's full attention, as is often the case with advertising.

I propose that we cannot use this type of data to make comprehensive claims about all types of multimodal texts and all instances of multimodal meaning construction. In our daily lives we face a wealth of word–image combinations every day. Due to the sheer production

volume, the reality of multimodal text creation is that the modes may not always be meticulously orchestrated to work together in the best possible way: the operating manual of a new model of a device could go into print with images of the previous model, a news article in fast-paced online reporting could be published with a completely wrong photo, and so on. As this paper aims to demonstrate further on, even if the modes are not designed into a coherent whole, they may still interact, that is, the reader may still interpret them in relation to each other. Can we assume that the same word–image relationships hold for illustrated texts in which the modes are less harmonized?

I also wish to raise two other points related to one of the main arguments of the paper, namely that the process of multimodal meaning construction is highly unpredictable. Much of the research theorizing about the way we inspect images concentrates on describing, for instance, the relative size, colours or tonal contrast of the elements of an image, the assumed scanning path of left to right, and so on. Obvious examples of such research are Kress and van Leeuwen's *Reading Images* (1996) and the abundance of studies building on their ideas. However, viewing images is a task-oriented activity. Our gaze is directed by our individual preferences, goals and expectations, emotions and prior knowledge (e.g. Boeriis and Holsanova 2012:262). In fact, a growing body of research within studies of visual cognition suggests that when viewing images, our gaze is directed more by our cognitive information-gathering needs than it is by the inherent visual properties of the image (for reviews on research, see e.g. Mills et al. 2011; Henderson et al. 2007).

What I wish to emphasize is that we cannot conclusively predict the way in which an image will be interpreted by individual readers by examining the properties of the image only. Different viewers may attribute different meanings to the same image, and even the same viewer may attribute different meanings to the same image when examining it for a different purpose. Similar claims can obviously be made about the way we interpret verbal texts. The idea that each reader interprets verbal text in an individual way has been around for decades in various lines of research, such as reader response criticism (e.g. Rosenblatt 1978), audience reception theory (Hall 1980), cognitive linguistics (e.g. Langacker 1991) and linguistic anthropology (e.g. Ottenheimer 2013). This idea is also present in Translation Studies (e.g. Rydning and Lachaud 2010). If we cannot predetermine the way in which words and images will be interpreted in isolation, we should be very careful in making claims about the way in which they will be interpreted when presented together.

The second point I wish to consider is whether (or not) we may assume that the reader always interprets the modes in relation to each other – that the reader even acknowledges both modes with equal attention. This seems to be taken as a premise in much of the research theorizing about word–image interaction, but it is a premise that needs to be verified. Evidently, if we analyze word–image interaction in a page-wide ad from a magazine, it is reasonably safe to assume that the reader will inspect and interpret both modes. It is also quite safe to make similar assumptions about the reading situation of picturebooks or graphic novels. Yet, one may ask if this is an assumption that will always hold.

The data analyzed in the present research project, described in the following section of the paper, introduces an illustrated text in which words and images offer differing and even contradictory perspectives into their shared referent – interaction we could call asymmetry without narrative function. Examining such data allows us to see that the proposed categories of word–image relationships, both the taxonomies as well as the grammatically-based networks, fall short in their ability to describe the different ways in which the reader might combine the information conveyed by the modes. The variety of interpretation displayed in the example sentences highlights that the interaction of modes in a particular word–image pair can often not be described by giving it a label (for instance, *complementing* or *developing*).

### **3. Empirical Insights from the Research Project**

The research project examined how a group of eight Master's level translation students translated a multimodal source text from English to Finnish, produced at the University of Tampere, Finland, as a part of a technical translation course. All of the MA students spoke Finnish as their native language, and had participated on basic and intermediate English translation courses prior to participating on the course. It is reasonable to suggest that translations produced by translation students might differ from the translations produced by experienced professionals. Yet, as Sager (1994:152) proposes, one of the advantages of conducting research on translation students is that they represent the diversity of human translators, which Sager describes as the most important variable in translation.

The source text used for the translation assignment presented the illustrated operating principles of two types of ore beneficiation devices used in the mining industry, namely concurrent and counter-current magnetic separators. Selecting the subject domain for the source text was based on an attempt to maximize the students' involvement with the images:

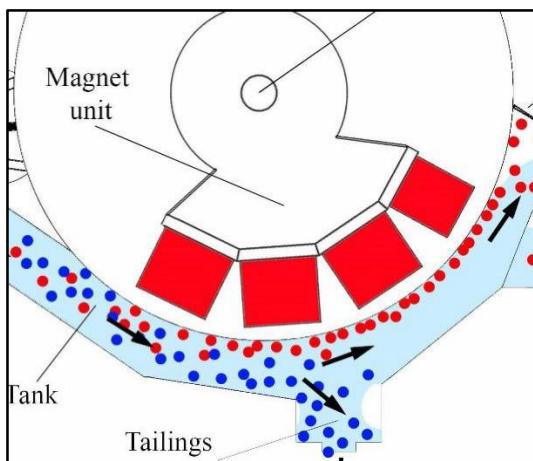
research on illustrated text comprehension has shown that the less the readers have prior knowledge of the subject domain, the more they resort to the images for their information gathering needs (Mayer and Gallini 1990:724; Hegarty and Just 1993:736). A background information questionnaire confirmed that the students did not consider themselves familiar with the topic: four of the students assessed their background knowledge as very poor, four as poor. The source text consisted of roughly 500 words and two large colored images, positioned along two pages in close proximity to each other. As mentioned above, in some parts of the illustrated text, the information conveyed by the two modes was, in one way or another, asymmetrical. These instances of asymmetry are discussed in more detail in the examples below.

The study examined the translation process from two different perspectives, employing two sets of data: examining the translations (see Ketola 2016a) as well as analyzing the translation diaries produced during the assignment (see Ketola 2015). Translation diaries are a form of introspective reporting conducted during and after the translation assignment. In these reports, the translators reflect on the problems they encountered during translation, how they solved them and why. A week after the translation assignment, the translators were gathered to inform them about the aims of the study, to ask for research permissions and to give them feedback on the translation assignment. During this meeting, one of the translators reported having looked at the images before starting translation, but having disregarded them completely while doing the actual translation assignment. As discussed below, this piece of information became useful in the analysis of the empirical data.

This section presents four examples that are relevant for the argument of the paper (for more detailed discussion on the actual translation solutions and additional examples, see Ketola 2016a). Each of the examples examines the translation of a particular verbal element. I first introduce the complete phrase from which the element is extracted (with the analyzed element in bold), followed by a close-up from the accompanying image representing the corresponding part of the device or the operating process. Examples of Finnish translations are followed by English back-translations, made by the author of the paper.

**Example 1.** *The magnetic particles are separated from the rest of the stream as they adhere to the drum surface in the area of **the magnet**.*

Figure 1: Source-text image depicting the magnet unit.



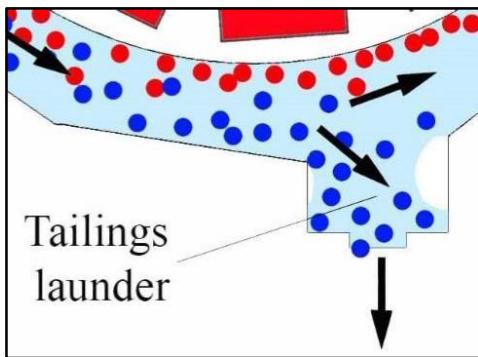
The first example deals with how the translators constructed an interpretation of the word *magnet* when presented together with the above visual information. The word refers to the magnet unit as a whole and is therefore in the singular form. Yet, in the corresponding part of the image, the magnet unit is depicted as consisting of four individual magnets (four large, red rectangles in Figure 1). As proposed above, examining how the group of translators conveyed the word *magnet* into the target language reflects how they constructed an interpretation from the combination of the modes. One of the translators omitted the prepositional phrase *in the area of the magnet* altogether, but the seven who preserved it all conveyed the word *magnet* into the target language in the plural form. Their translations for *in the area of the magnet* included solutions such as *magneettien vaikutusalueella* ('in the area of influence of the magnets') and *magneettien kohdalla* ('in the area of the magnets'). Finnish nouns are of the countable type and have different forms for singular and plural. There is no grammatical reason why the noun *magnet* should appear in plural in Finnish; the singular form would also be appropriate in the given context, and it would indeed better correspond to the original idea in the verbal source text. It hence seems that the translators examined both of the modes and that the image affected the way in which they comprehended the verbal element.

The relationship between these verbal and visual details, as perceived by the translators, seems to fit in well in the classifications of word–image interaction introduced above. In Martinec and Salway's terms, this instance of multimodal meaning construction could be described as *elaboration*: the visual mode adds information to the verbal, making its meaning more specific. In Marsh and White's classification this relationship could be described as *developing*, since the visual mode specifies the meaning of the verbal. When examining this instance of word–image interaction from the perspective of the interpretations of these seven

individuals, the relationship of the modes could hence be described as either elaborating or developing.

**Example 2.** *The weakly magnetic and non-magnetic particles are carried forward by the stream and eventually discharged from a tailings **launder** in the upper part of the equipment.*

Figure 2: Source-text image depicting a launder.



The second example deals with the translation of the word *launder*. There are three separate launders in the separators introduced in the source text which means the term is repeated throughout the text. In the mining industry, the term typically refers to a trough or a long narrow channel used to convey liquids in a vertical direction. Yet, in the source text images, the shape of the launders does not correspond to their verbal definition: they are depicted as slightly wider than they are tall (see, for instance, the tailings launder in Figure 2).

The group of translators produced four conclusively different translation solutions for *launder*. Only one of them produced a translation solution (*putki*, ‘pipe’) that describes a trough which could convey liquid in a vertical direction. During the post-translation meeting with the students, it was confirmed that this solution was made by the translator who did not employ visual information during translation. From the perspective of this translator, then, multimodal integration of meaning did not take place, and the essence of the message, as conveyed to the target language, consisted of verbal information only.

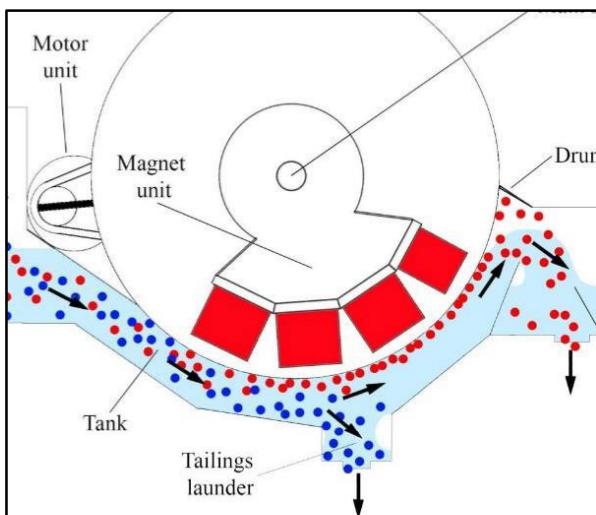
The remaining seven translators produced solutions that diverged from the verbal original in one way or another. I will here discuss the most commonly employed translation solution, *säiliö* (‘container’), used by four translators (please refer to Ketola 2016a for the rest). The translation solution is problematic. A container is used to *contain* something; it does not refer to a channel through which liquid passes. In the translation diaries, one of these four translators affirmed that the translation solution was indeed created in keeping with the shape

of the part in the image, describing how ‘container’ seemed like an apt choice of words ‘since it is so generic it can be of any shape at all’ (Ketola 2015:28). When deciding the best possible way of conveying this information to the target language, it appeared to be more important to make the solution match the visual depiction of the part rather than its verbal description.

The translation (‘container’) reflects how the verbal element (‘launder’) was interpreted in relation to the image. This instance of meaning construction does not adhere to any of the word–image relationships proposed by Martinec and Salway: nothing is being projected and nothing is being added. The meaning of the verbal element cannot be claimed to become more specific after being coupled with visual information. Neither can this particular instance of word–image interaction be described by any of the 49 relationships proposed by Marsh and White which also only account for words and images offering compatible perspectives into their shared referent. Their classification does introduce a concept of *contrast*, but it is used to describe instances of words and images opposing each other for rhetorical purposes – which is clearly not the case here. If we compare the verbal element with its translation, we could suggest that the visual information *distorts* the verbal in a way. Yet, I emphasize that this interpretation only applied for four of the translators.

**Example 3.** *The weakly magnetic and non-magnetic particles are carried forward by the stream and eventually discharged from a tailings launder in the upper part of the equipment.*

Figure 3: Source-text image depicting the location of the tailings launder.



Example 3 discusses an instance of contradiction between visual and verbal information: the location of *the tailings launder* is represented differently in the two modes. According to the verbal source text presented in Example 3, the tailings launder is located in *the upper part* of the equipment. Yet, in the image (Figure 3), the launder is one of the bottommost parts of the device. Examining how the translators constructed their interpretation of the location of the launder was one of the most interesting parts of the data. For six of the translators, this negotiation led to conveying the information offered by the verbal source text; two of them conveyed the information offered by the image. Most translators (all seven who actually employed visual information during translation) discussed this contradiction of information in their translation diaries, which means that they acknowledged both modes with enough attention to notice the contradiction.

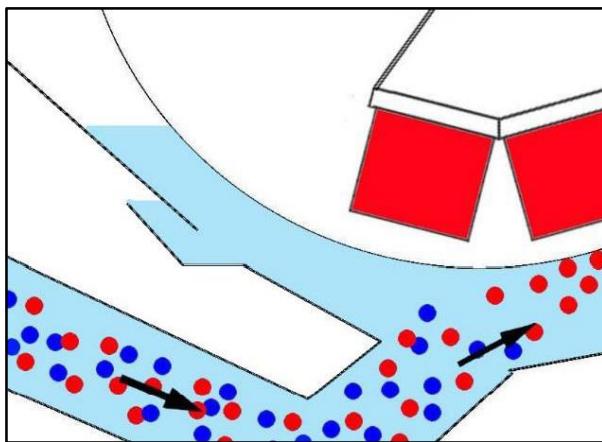
Again, the classifications introduced above do not offer a concept that could be used to describe this relationship. In principle, one could claim that if verbal and visual information so obviously contradict each other, we could simply label this as a mistake in the subject matter and refrain from defining a relationship between the modes. However, as the data of study readily displays, a negotiation of meaning between the modes has obviously taken place. From the perspective of the translator, the modes have interacted, and interaction supposes a relationship.

However, the fact that the modes provide contradictory information will not always result in translators constructing their interpretation by choosing between one or the other: the translator does not necessarily inspect both modes thoroughly enough to acknowledge a contradiction of information. Example 3 reflects translators choosing between one mode and the other when the modes are contradictory to each other. Yet, the data also included another example of contradictory information that did not prompt a negotiation between the modes for most of the translators. This example regarded a part of the device referred to as the *drum*. According to the verbal text, the drum was ‘submerged in a tank’ – entirely sunk below the surface of water – but, in the image, only the bottommost part of the drum was under water. In the translation diaries, only two of the translators discussed this contradiction, one of whom decided to change the wording to a more general expression so that contradiction between word and image was eliminated. In the post-translation meeting, the other six affirmed not having noticed the contradiction. This illustrates that the process of multimodal meaning construction is unpredictable: for most of the translators, the modes did not interact in this part of the source text.

Anne Ketola, *Using Translation Research to Model Word–Image Interaction*, 82 – 104.

**Example 4.** *The non-magnetic particles flow in the opposite direction to the drum rotation and are discharged through **an overflow** into a tailings chute.*

Figure 4: Source-text image in which the exit route of the non-magnetic particles is not depicted.



Example 4 illustrates how the lack of visual information may affect the translator's verbal text comprehension. The example represents a part in the source text describing how the stream of non-magnetic particles separates from the stream of magnetic particles in a counter-current separator. However, the corresponding information cannot be found in the image (Figure 4): while the stream of the magnetic particles (red dots) continues its journey to the right, the stream of the non-magnetic particles (blue dots) would continue its course to the left. Yet, the non-magnetic particles are not depicted in the image beyond the point where they separate from the magnetic ones. The *overflow* or the outlet through which the particles then exit the separator can be seen in the image, but the name of the part of the equipment is not verbally labeled in the image (cf. verbal labels in Figures 1, 2 and 3). Further, the *tailings chute* through which the particles continue their journey after exiting the device (mentioned in the end of the example sentence) is not depicted in the image at all.

Examining how the term *overflow* was interpreted by the translators exemplifies how the lack of visual information may affect the translators' multimodal text comprehension. *Overflow* can refer to both the flowing over of a liquid (either as a noun or a verb) or, as in this particular text, an outlet for excess liquid. Only two translators produced a solution which conveyed the idea of *an overflow* as a passage through which liquid is let out. The rest of the translation solutions reflected a somewhat mistaken idea of the element, describing *overflow* Anne Ketola, *Using Translation Research to Model Word–Image Interaction*, 82 – 104.

as flow of liquid, either as a noun or as a verb (e.g. ‘discharged as overflow’, ‘by overflowing’). If we compare these translation solutions to the original verbal element, we could again claim that the original meaning has been somewhat distorted in the translation process (cf. Example 2).

All in all, the analysis of the translation solutions strongly implied that the translators struggled to understand what the word referred to exactly in this particular text. One translator omitted the prepositional phrase altogether, which can also be indicative of not understanding the phrase. The translation diaries confirmed that this indeed was the case: most of the translators explicitly reported not having understood the phrase, even though the phrase in itself was not more complicated than the rest. In Ketola (2015:27) I suggested that the fact that the translators struggled to comprehend the phrase reflects a strong effort to negotiate the modes in relation to each other. If verbally acquired information cannot be confirmed from the image, the translator may be left with an impression of not having understood the verbal information at all.

It is quite understandable that none of the word–image relationships proposed in the classifications seem to cover this example. If visual information is indeed missing in this fashion, it is easy to claim that there is no word–image interaction; that we cannot describe interaction between verbal information and something that does not exist. It is thought-provoking, however, to notice that the translators displayed such a vigorous attempt to interpret verbal information in relation to the visual that the lack of visual information actually disturbed their verbal text comprehension. We will never arrive at observations such as these if we limit our analyses to word–image pairs which represent the typically examined text types described above. Analyses based on limited data offer limited possibilities for observations.

To conclude, the analysis of the data reflected the manifold, unpredictable nature of multimodal meaning construction. Even with a limited sample size of eight translators, variety in the interpretation was evident. Moreover, the analysis of the data illustrated that the reviewed classifications of word–image relationships, both the taxonomy as well as the grammatically-based network, do not cover all of the relationships that may hold between verbal and visual information.

It is reasonable to argue that a group of experienced translators would have translated the source text differently than a group of students (cf. e.g. Göpferich 2010). Further, one could argue that a group of experienced translators would have resorted to the images of the source text less than the students did: As discussed in the beginning of this section, the less the readers have prior knowledge of the subject domain of the text, the more likely they are to resort to the images for information gathering purposes. It is therefore likely that a group of experienced technical translators might have relied on the images less than the students did and that, consequently, their translation solutions would not have reflected the visual dimension of the source text as much as the students' solutions did. Yet, I propose that, as to the contribution of my study to multimodality research, the importance of the observations is not affected by the student status of the research subjects. In the quest to understand the complexity of how we interpret the combination of words and images, it is equally important to examine readers with high and low background knowledge of the subject domain of the illustrated text in question.

#### **4. Conclusions**

The paper set out to discuss multimodal meaning construction both in Translation Studies as well as on a more general level. From the perspective of Translation Studies, the discussion presented in the paper has aimed to emphasize that when translating multimodal material, all of the modes of the source text can be involved in the translator's interpretation of the source text. Much of the material that is being translated today is multimodal (Hirvonen and Tiittula 2010:1). The rapidly developing multimodal text-production practices also urge the discipline to continue shifting its focus away from linguistics towards multimodally-informed perspectives.

As to research into multimodality on an interdisciplinary level, the paper has aimed to demonstrate that translation analyses may offer one possible means to empirically exemplify multimodal meaning construction, and that based on the analyses conducted in my previous research, the way in which word–image interaction has been previously theorized does not always seem to hold. I have reviewed two ways of classifying possible relationships that may hold between words and images when the two modes are presented together, which are based on widely acknowledged research. I have suggested that the relationships proposed in the reviewed classifications do have a value if used as concepts describing individual interpretations in empirical research (see Example 1 above). They also have a value if used as

hypotheses for multimodal meaning construction or as concepts describing the overall meaning-constructing potential of the modes. However, based on my translation analyses, I have argued that these classifications need to be employed with caution for two reasons.

First, we must acknowledge that, despite their apparent comprehensiveness, such classifications cannot not be exhaustive. The empirical analysis displayed ways of interpreting the combination of verbal and visual information that were not represented in the classifications. Depending on ready-made classifications in our analyses of word–image interaction can therefore obscure the fact that other relationships may also hold between the two modes. A possible line of future research into word–image interaction could entail complementing the classifications with research performed on more versatile data. The list of possible relationships, however, might turn out to be endless, as it represents the diversity of the human cognitive repertoire. I therefore argue that pursuing ‘a complete set’ of possible relationships between words and images might be a counterproductive effort and do very little to advance our understanding of how the combination of the modes is actually perceived by the readers of multimodal artifacts.

My second main argument is that suggesting a relationship between a word–image pair does not allow us to *predict* the way in which the modes will interact from the perspective of the reader when presented together. This interaction, as perceived by a reader, is an individual experience. Readers interpret the combination of the modes in various ways, depending on factors such as their cognitive information-gathering needs, individual preferences and prior knowledge. Moreover, from the perspective of the individual reader, the modes do not always interact even if various modes are simultaneously presented for the reader. As the analysis of the examples showed, the reader does not necessarily acknowledge both modes with equal attention at all times.

Word–image interaction has received great research interest in a variety of disciplines; yet, this research tends to be empirically unsubstantiated (cf. Bateman, quoted in the Introduction). In this paper, I have proposed that research into translation can offer one possible way to empirically exemplify how the combination of words and images is interpreted. Yet, research into multimodality faces an urgent need for further empirical work from, preferably from various methodological positions. Theoretical frameworks describing word–image interaction are abundant. We now need to test these ideas in ways that go beyond the authors’ own interpretations.

## Author details

Anne Ketola (PhD student)  
University of Tampere, Faculty of Communication Sciences,  
Kanslerinrinne 1, Pinni B, Room B5044,  
33100 Tampere, Finland.  
Tel. 00358 504902875  
Email: [anne.ketola@uta.fi](mailto:anne.ketola@uta.fi)

## References

- Barthes, Roland (1964) ‘The rhetoric of the image’, in Roland Barthes (ed.) *Image-Music-Text*, London: Fontana, 32-51.
- Bateman, John A. (2014) *Text and image: A critical introduction to the visual/verbal divide*, London: Routledge.
- Boeriis, Morten and Jana Holsanova (2012) ‘Tracking visual segmentation. Connecting semiotic and cognitive perspectives’, *Visual Communication*, 11(3): 259-281.
- Connors, Sean P. (2013) ‘Weaving multimodal meaning in a graphic novel reading group’, *Visual Communication*, 12(1): 27-53.
- Göpferich, Susanne. (2010) ‘The Translation of Instructive Texts from a Cognitive Perspective: Novices and Professionals Compared’, in Susanne Göpferich, Fabio Alves and Inger M. Mees (eds.) *New Approaches in Translation Process Research*, Copenhagen: Samfundsletteratur, 5–55.
- Hall, Stuart (1980) ‘Encoding/Decoding’, in Stuart Hall, Dooty Hobson, Andrew Lowe and Paul Willis (eds.) *Culture, Media, Language: Working Papers in Cultural Studies 1972–79*, London: Hutchinson, 128-138.
- Halliday, M. A. K. (1985) *An Introduction to Functional Grammar*, London: Edward Arnold.
- Hegarty, Mary, and Marcel A. Just (1993) ‘Constructing Mental Models of Machines from Text and Diagrams’, *Journal of Memory and Language*, 32(6): 717-742.
- Henderson, John M., James R. Brockmole, Monica S. Castelhano and Michael Mach (2007) ‘Visual Saliency Does Not Account for Eye Movements During Visual Search’, in Roger P.G van Gompel, Martin H. Fischer, Wayne S. Murray and Robin L. Hill (eds) *Eye Movements: A Window on Mind and Brain*, Amsterdam: Elsevier, 537-562.
- Anne Ketola, *Using Translation Research to Model Word–Image Interaction*, 82 – 104.

- Hirvonen, Maija and Liisa Tiittula (2010) ‘A method for analysing multimodal research material: Audio description in focus’, *MikaEL – Electronic Journal of the KäTu Symposium on Translation and Interpreting*, 4: 1-14.
- Holsanova, Jana (2012) ‘New methods for studying visual communication and multimodal integration’, *Visual Communication*, 11(3): 251-257.
- Ketola, Anne (2015) ‘Translation diaries of an illustrated technical text. Translation students’ conceptions of word-image interaction’, *Connexions Special issue: Translation and international professional communication: Building bridges and strengthening skills*, 3(2): 13-40.
- Ketola, Anne (2016a) ‘An Illustrated Technical Text in Translation. Choice Network Analysis as a Tool for Depicting Word-Image Interaction’, *trans-kom*, 9(1): 79-97.
- Ketola, Anne (2016b) ‘Towards a multimodally oriented theory of translation: A cognitive framework for the translation of illustrated technical texts’, *Translation Studies*, 9(1), 67-81.
- Kong, Kenneth C. C. (2006) ‘A taxonomy of the discourse relations between words and visuals’, *Information Design Journal*, 14(3), 207-230.
- Kress, Gunther and Theo van Leeuwen (1996) *Reading Images. The Grammar of Visual Design*, London: Routledge.
- Kress, Gunther and Theo van Leeuwen (1998) ‘Front pages: the critical analysis of newspaper layout’, in Allan Bell and Peter Garrett (eds) *Approaches to Media Discourse*, Malden: Blackwell Publishers, 186-282.
- Kussmaul, Paul (2000) ‘A Cognitive Framework for Looking at Creative Mental Processes’, in Maeve Olohan (ed.) *Intercultural Faultlines. Research Models in Translation Studies I. Textual and Cognitive Aspects*, Manchester: St. Jerome, 57-72.
- Langacker, Ronald W. (1991) *Foundations of Cognitive Grammar. Descriptive Application*, Stanford: Stanford University Press.
- Levin, Joel R. and Richard E. Mayer (1993) ‘Understanding illustrations in text’, in Bruce K. Britton, Arthur Woodward and Marilyn R. Binkley (eds) *Learning from Textbooks: Theory and Practice*, Hillsdale: Lawrence Erlbaum Associates, 95-113.

- Liu, Yu & Kay O'Halloran (2009) 'Intersemiotic Texture: analyzing cohesive devices between language and images', *Social Semiotics*, 19(4), 367-388.
- Marsh, Emily E. and Marilyn Domas White (2003) 'A taxonomy of relationships between images and text', *Journal of Documentation*, 59(6): 647-672.
- Martinec, Radan and Andrew Salway (2005) 'A system for image-text relations in new (and old) media', *Visual Communication*, 4(3): 339-374.
- Matthiessen, Christian (2007) 'The Multimodal Page: a Systemic Functional Exploration', in Terry D. Royce and Wendy L. Bowcher (eds.) *New Directions in the Analysis of Multimodal Discourse*, Mahwah: Lawrence Erlbaum Associates, 1-62.
- Mayer, Richard E., and Joan K. Gallini (1990) 'When is an illustration worth ten thousand words?', *Journal of Educational Psychology*, 82(4): 715-726.
- Mills, Mark, Andrew Hollingworth, Stefan Van der Stigchel, Lesa Hoffman and Michael D. Dodd (2011) 'Examining the influence of task set on eye movements and fixations', *Journal of Vision*, 11(8): 1-15.
- Moya Guijarro, Arsenio Jesus (2014) *A Multimodal Analysis of Picture Books for Children. A Systemic Functional Approach*, Sheffield: Equinox.
- Nikolajeva, Maria and Carole Scott (2000) 'The dynamics of picturebook communication', *Children's Literature in Education*, 31(4): 225-239.
- Ottenheimer, Harriet Joseph. 2013. *The Anthropology of Language: An Introduction to Linguistic Anthropology*. 3rd ed. Belmont: Wadsworth.
- Rosenblatt, Louise M. (1978) *The Reader, the Text, the Poem: The Transactional Theory of the Literary Work*, Carbondale: Southern Illinois University Press.
- Rydning, Antin Fougner and Christian Lachaud (2010) 'The reformulation challenge in translation context reduces polysemy during comprehension, but multiplies creativity during production', in Gregory M. Shreve and Erik Angelone (eds) *Translation and Cognition*, Amsterdam: John Benjamins, 85-108.
- Sager, Juan C. (1994) *Language engineering and translation: Consequences of automation*, Amsterdam: John Benjamins.
- Spillner, Bernd (1982) 'Stilanalyse semiotisch komplexer Texte', *Kodikas/Code Ars Semeiotica*, 4/5(1): 91-106.
- Anne Ketola, *Using Translation Research to Model Word–Image Interaction*, 82 – 104.

Unsworth, Len and Chris Cléirigh (2009) ‘Multimodality and reading: The construction of meaning through image-text interaction’, in Carey Jewitt (ed.) *The Routledge Handbook of Multimodal Analysis*, London: Routledge, 151-163.