

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 Martín (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

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

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)¹.

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–

¹ 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:

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

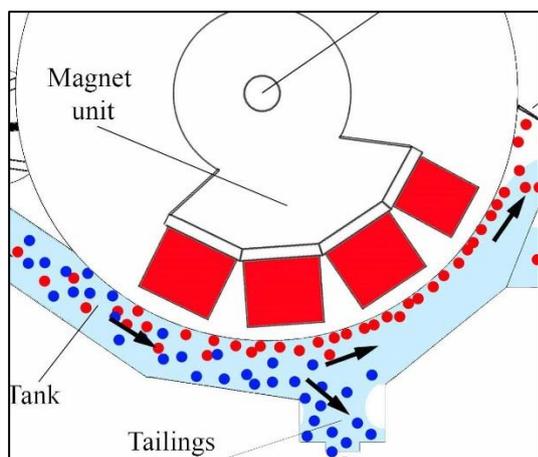
research on illustrated text comprehension has shown that the less the readers have prior knowledge of the subject domain, the more they resort to the images for their information gathering needs (Mayer and Gallini 1990:724; Hegarty and Just 1993:736). A background information questionnaire confirmed 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.



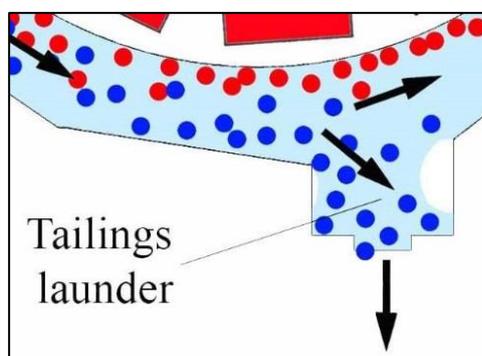
The first example deals with how the translators constructed an interpretation of the word *magnet* when presented together with the above visual information. The word refers to the magnet unit as a whole and is therefore in the singular form. Yet, in the corresponding part of the image, the magnet unit 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.

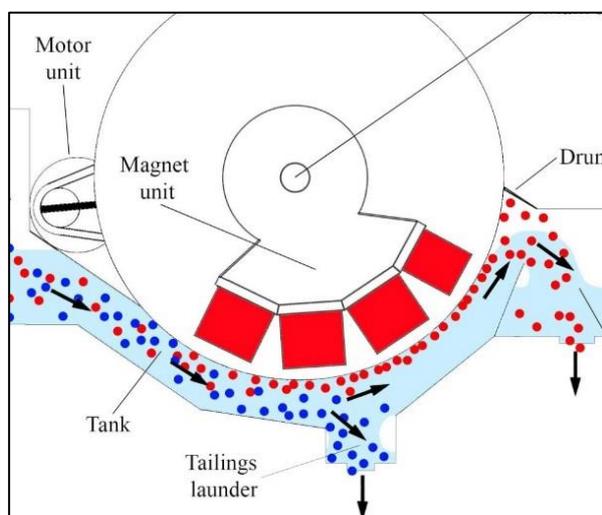
The remaining seven translators produced solutions that diverged from the verbal original in one way or another. I will here discuss the most commonly employed translation solution, *säiliö* (‘container’), used by four translators (please refer to Ketola 2016a for the rest). The translation solution 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.



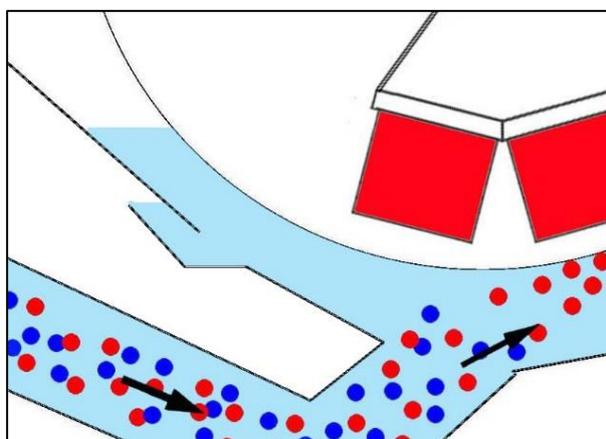
Example 3 discusses an instance of contradiction between visual and verbal information: the location of *the tailings launder* is represented differently in the two modes. According to the verbal source text presented in Example 3, the tailings launder is located in *the upper part* of the equipment. 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.

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

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: Samfundslitteratur, 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. Castelhana 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.