24

Translation, multimodality and cognition

Jan-Louis Kruger

24.1 Introduction and background

There have been a number of publications on translation and multimodality over the past decade (cf. e.g. de Pedro Rico, 2012; González, 2014; Kaindl, 2013; O’Sullivan, 2013; Tuominen et al., 2018). On the whole, however, Translation Studies has not awarded much attention to multimodality, and even less to multimodality and cognition. When considering source and target texts, the emphasis traditionally has been on the linguistic level. Kaindl (2013, p. 257) points out that both diachronic and synchronic Translation Studies in the past focused mainly on the linguistic dimension, rendering the discipline monomodal.

This emphasis on language rather than multimodal meaning transfer has been a blind spot in translation research for much of the history of the discipline. But then, as Gambier (2006:6) points out, “[n]o text is, strictly speaking, monomodal. Traditional texts, hypertexts, screen texts combine different semiotic resources. Films and TV programs co-deploy gesture, gaze, movement, visual images, sound, colours, proxemics, oral and written language, and so on”. And as Tuominen et al. (2018, p. 1) point out, most communication in contemporary society has become primarily multimodal, and this multimodality is therefore also a central concern in translation and gives rise to interesting questions about the optimal methodology to study multimodality in translation. In fact, according to Tuominen et al. (2018, p. 5), we could refer to multimodal Translation Studies as “a translation-specific derivation from the term ‘multimodal studies’, to foreground multimodally oriented translation research as a research orientation with a specific objective of addressing the presence and interplay of different modes in translational contexts”.

The preceding reference to different text types also reveals the skewed emphasis in Translation Studies on the translation product. It is, however, what happens to this product when the target text is received and processed by the target-language audience, as well as how it comes into being (in other words, the production thereof), that makes it imperative to consider cognition.

Multimodality is a key concern in interlingual interpreting, where the interpreter has to provide not only a translation of the words of the source-language speaker but also an interpretation of how these words are presented (e.g. intonation) as well as of visual information that forms part of the context of what is spoken (e.g. gestures, notes, printed matter or slides). Sign language interpreting likewise mediates between an auditory spoken language (with a similar
multimodal context as in interlingual interpreting) and a deaf audience who rely on a visual interpretation of the spoken language through a complex set of signs, including facial and bodily expressions. In audiovisual translation, the linguistic modes of spoken and written language (in signs and subtitles) always co-exist with various other modes such as moving and static images, film editing, camera angles, mise-en-scène, music and other sounds, presented in the medium of film, whether - in a theatre or digital on television or on the Internet. Even the translation of written text often has to consider extra-linguistic aspects such as typeface, layout and various forms of illustration or graphics that go beyond the single modality of written language.

Translation Studies has embraced cognitive approaches over the past decades, and indeed, according to Muñoz Martín, “cognitive approaches to translation and interpreting may be considered the oldest empirical research area of modern translation studies” (2016, p. 555). Cognitive Translation Studies deals with the processes in the production and reception of translation and interpreting products in an interdisciplinary manner, focusing on the cognitive processing of translators and interpreters as well as the users of translation and interpreting. In that sense, there is a strong link between multimodality and cognition in translation that becomes evident in empirical studies where ergonomics, the interaction between the translator and various modes and sources of information in different media, is investigated. Multimodality is also central in studies on the difference in processing between reading and writing, as well as between watching, listening and speaking. This emphasis on multimodality is evident in a number of recent books on Cognitive Translation Studies and translation process research (see Lacruz & Jääskeläinen, 2018; Muñoz Martín, 2017; Schwieter & Ferreira, 2017; Shreve & Angelone, 2010).

24.2 Multimodality

Before looking at cognition in the translation of multimodal texts, it might be useful to clarify what multimodality means. This is by no means a clear concept, but there seems to be some convergence on meaning in work on multimodality in the context of social semiotics (cf. Halliday, 1978). In this regard, it is important to distinguish between mode and medium. According to Kress and van Leeuwen (2001), mode is related to resources for making meaning that employ channels of communication or representation. It therefore includes linguistic elements, such as spoken dialogue and written text, as well as static and dynamic images, sound, gesture, gaze, facial and bodily expressions or posture. This conceptualization of mode and multimodality therefore takes the emphasis away from simply linguistic components. Medium, on the other hand, relates to the material form (see Littau, 2011) that conveys the message contained in the mode, such as paper, ink, film or hyperspace. Kress and van Leeuwen (2001) argue that medium has traditionally been neglected in both linguistics and semiotics, just as non-verbal modes have been neglected.

Fundamentally, however, looking at translation from a multimodal perspective poses a number of questions about the very nature and self-image of a discipline that has been conceptualized primarily on the basis of monomodality and linguistic texts and speech units (see Kaindl, 2013, p. 257). This is not to say that Translation Studies has been focused exclusively on the linguistic aspects of “monomodal” texts. Translation Studies has, in fact, been concerned with systems and agents that shape the translation product for many years. This is evident in Jakobson’s (1959/2000) identification of intersemiotic translation as a category of translation; in Holmes’s concept of function-oriented descriptive Translation Studies (or DTS), which is concerned with context rather than text (Holmes, 1988/2000); in polysystems theory, which considers literary translation within a broader literary system (Even-Zohar, 1978/2000); and in Reiss’s (1981/2000) addition of audiomedial as a text type in addition to informative, expressive and operative texts, among
others. This broader conceptualization of multimodality foregrounds the fact that, regardless of our definition of text, the linguistic content is always framed in, supplemented by and informed by context, co-texts and non-verbal modes that determine the translation as well as the reception of the translated product (see, for example, Kruger & Kruger, 2017).

Interestingly, when defining multimodal discourse in their theory of multimodal communication, Kress and van Leeuwen (2001, p. 21) saw modes as “semiotic resources which allow the simultaneous realization of discourses and types of (inter)action. Designs then use these resources, combining semiotic modes, and selecting from the options which they make available according to the interests of a particular communication situation”. Kaindl (2013, p. 258) interprets this as meaning that modes are not in the first instance products, “but cultural processes which manifest themselves as discourses and the functions of which constitute texts in relation to other modes”.

Klaus Kaindl’s lucid and comprehensive contribution on “Multimodality in translation studies” in The Routledge handbook of translation studies (2013) provides an excellent critical overview of the concept of multimodality and how this has been approached in Translation Studies. Many of the problems in these approaches can be related to the confusion between mode and medium, and the use of semiotic systems in defining multimodality.

Toury (1994) defines intersemiotic translation as the translation between different codes, whereas interlingual translation is considered to be intrasemiotic translation (divided into intrasystemic or intralingual and intersystemic translations). Kaindl (2013, p. 261) therefore argues that the criteria of mode and medium should be clearly distinguished from each other and that the semiotic dimension is problematic for a translation-relevant text typologization such as those presented by Jakobson and Toury.

Having identified mode and medium as the categories that have to be distinguished in translation, Kaindl (2013, pp. 261–262) suggests a distinction between intramodal and intermodal translation in the dimension of mode or semiotic code, and between intramedial and intermedial translation (or transfer aspects) in the dimension of medium referring to the materialities of translation. In intramodal translation, one mode is translated with the same form of mode (i.e. linguistic to linguistic, such as in translation between dialects of the same language in a drama; image to image, such as in a translation of Disney cartoons into manga with an emphasis on the styles of animation; or one musical genre into another, etc.). Intermodal translation involves a change in mode, such as a translation from a linguistic mode to an image mode or vice versa (as in audio description). Intramedial translation maintains the same medium while transferring media elements in accordance with cultural conventions related to the medium (i.e. adhering to medium-related conventions in drama for different cultures in translation). Intermedial translation, by contrast, concerns translation across media barriers (such as the translation of a novel into a film or a play into a musical). Kaindl (2013) draws attention to a range of areas in Translation Studies that have engaged with multimodality (such as audiovisual translation, the translation of children’s literature, the translation of pragmatic texts like advertisements, the localization of websites or video games). He also points out the lack of attention to multimodality in the areas of specialized texts (such as technical communication) and interpreting (where surprisingly little work has been done on multimodality up until recently—see Kaindl, 2013, p. 265 for some exceptions). Kaindl (2013, pp. 264–265) concludes by identifying a need for an extension not only of the definition of text as basis for translation to include multimodality, but also for an extension of the instruments used for analysis of these texts in Translation Studies. This centres primarily on text analysis, where he distinguishes between, firstly, an analysis of the composition and functioning of different modes, and, secondly, the correlation and interaction modalities between different modes beyond the verbal modes.
In the short period since Kaindl’s seminal chapter in 2013, multimodality in translation has received significantly more attention (see, for example, O’Sullivan, 2013; Tuominen et al., 2018), particularly in audiovisual translation and in translation process research. In the field of audiovisual translation, for example, the increased focus on audio description has seen a number of studies and large projects aimed at investigating multimodality in translating audiovisual texts and contexts. In translation process research, multimodality has taken centre stage in the investigation of the way in which translators and interpreters process source texts and contexts and produce target texts, and significantly, on how audiences process these target texts and contexts. This has necessarily also seen a growing emphasis on the investigation of cognition. In the rest of this chapter, these areas will be discussed in more detail in order to show how multimodality and the cognitive processing of multimodality have shaped recent developments in Translation Studies.

24.3 Core issues in multimodality and cognition

Measuring the impact of multimodality on the production and processing of translation is an important step towards understanding how different modes impact on different aspects of cognition, such as knowledge, beliefs, attention, emotions and memory, in this complex environment. But what makes multimodality such a complicated issue to study? Part of the answer lies in the fact that the human mind has the capacity to process and integrate various different sources of information not only in sequence but also simultaneously. Multimodality means that we are not only forced to engage with the linguistic aspects of a multimodal text, but we also have to interrogate constantly how the different modes contribute to the creation of meaning. This is complicated by the fact that we process some elements of a multimodal text automatically: when text appears on a screen we automatically begin to read it (see d’Ydewalle and De Bruycker, 2007; d’Ydewalle and Gielen, 1992); when there is movement on a screen, our eyes are drawn to the movement; we tend to begin to process a video by looking at the centre of the screen or page. These automatic impulses are referred to as bottom-up processing and compete with top-down processing, where we consciously direct our attention to salient elements of a text.

The efficiency with which humans engage with multimodal contexts is obvious when we look at the astonishing speed with which infants pick up on cues around them and begin to interact with their surroundings. It is also evident in the way in which we can view a fiction film consisting of a series of distinct scenes and shots that are edited together, ignoring the discontinuities in these texts. We constantly pick up on overt and covert cues around us, including when we translate or process a translated text.

The relevance of all of this lies in the fact that multimodal texts consist of a number of modes that contribute to the creation of meaning. Much of this is redundant or interdependent. When we see an emotional scene in a film, we know a character is sad because of the words in the spoken dialogue, the intonation in the voices, the words in the subtitles, the expressions on faces, the musical score, the camera angles and close-up shots, and so forth (all of which, of course, also carry culturally defined meaning). In many cases, any one of these codes would be sufficient to give us the information we need. In that sense, the different sources could be considered redundant.

What makes redundancy such an interesting concept in the context of multimodality and cognition is that different users will process codes in a different order, or will process different codes to arrive at the same understanding. Redundancy has a number of functions. It often provides confirmation in one mode of what is presented in another. In that respect, it could be supplementary—adding more nuance to our understanding. In other cases, the redundancy...
could be partial, meaning that an interdependence will exist between two sources of information. But in some cases, redundancy could result in competition, introducing a complexity that forces the user to process a number of redundant sources of the same information in different modes simultaneously (see also Lautenbacher, 2018). Much of what we most need to understand, therefore, is how humans process multimodality cognitively.

In educational psychology and instructional design, the redundancy effect is associated with a negative impact of redundant information on comprehension and learning. This results in a potential overloading of the working memory that has a negative impact on performance (see Kalyuga, 2012; Moreno & Mayer, 2002). However, when there is an interdependence between information in verbal and visual modes, the redundancy effect does not occur—there is no dramatic increase in the cognitive load associated with processing both sources of information.

However, the impact of different sources of information varies from individual to individual. For this reason, it is essential to study the cognitive processing of multimodal texts—not only to provide us with a better understanding of how information presented in different modes is processed, but also to help us understand how translators code this multimodality. Audiovisual translation provides the ideal context for studying multimodality and the processing of various sources of information simultaneously.

**24.4 Audiovisual translation (AVT)**

In AVT, multimodality has long been acknowledged as a fundamental concern. Kaindl (2013) acknowledges that this field is at the cutting edge of the comprehensive and systematic description of “the role and function of non-verbal modes in transcultural communication” (2013, p. 263). In recent years, this focus has deepened (see e.g. Tuominen et al., 2018). For example, Taylor (2016) explores the multimodal approach in AVT by focusing on multimodal text analysis (including narrative, linguistic, semiotic and cultural considerations) before turning to multimodal transcription, which provides a useful methodology for engaging with the various modes that impact on AVT. He points out, however, that such transcription becomes too unwieldy when it comes to whole films, calling rather for phasal analyses that enable the translator “to identify homogeneous ‘phases,’ both continuous and discontinuous, within a multimodal text and to recognize register changes, character traits, and elements of cohesion and coherence that, if ignored, could lead to inconsistencies in translation” (2016, p. 230). This sense of economy in dealing with multimodality in AVT provides a scalable solution to an enterprise that could easily become bogged down in microtextual elements.

Audio description (AD) is the one form of AVT that engages with the multimodality of the text beyond the linguistic content of the dialogue to a much greater extent than any other form of AVT. Here, the dialogue is the one element of the text that is left undisturbed in the spoken form, while the various other modes that constitute film are mediated for the blind audience. Interestingly, much recent research on AD has centred on the difference between a focus on the unimodally visual component of what can be seen on screen in terms of characters, objects and action and a focus on the interaction between sound effects, visual presentation (in camera angles, editing and mise-en-scène) and verbal modes.

The foregrounding of multimodality in AD is also evident in a series of studies dealing with sound, film style and aesthetics. In a collection edited by Maszerowska, Matamala and Orero (2014), Elisa Perego (2014), for example, discusses the importance of visual composition and editing techniques for AD. This resonates with the work by Fryer and Freeman (2013), which also deals with multimodality in grappling with whether or not to include filmic language (descriptions of style) in AD. Mazur (2014) discusses interpretative description and shows that
the interpretation of the visual modes of gestures and facial expressions is sometimes essential in order to convey the filmic narrative as well as its aesthetic dimension. Matamala (2014) considers the transfer of visual verbal text to spoken word in describing text on screen, before Szarkowska and Orero (2014) discuss the importance of describing sounds in AD in order to disambiguate sounds for the target audience. These studies, while by no means the only ones engaging with multimodality in AVT, signal an important shift in this branch of Translation Studies away from the linguistic mode in isolation.

In the context of multimodal analysis, Jiménez Hurtado and Soler Gallego (2013) explore the possibility of applying corpus-based methods to the analysis of film for AD, which enables the narratological structure of the film, the filmic language and the linguistic elements of the film to be analysed in relation to the AD. This multimodality is also addressed by Wilken and Kruger (2016) in looking at the impact of filmic elements such as mise-en-scène (the way in which visual elements are shown to an audience) on the processing of the film and particularly the immersion of the audience in the film. The multimodal transcription of scenes from a film forms the basis for their analyses.

In addition to text analysis, the processing of multimodal texts by audiences has received ample attention in AVT. I will present a few examples from the fields of AD, subtitling and dubbing, with a distinction between the offline and online measurement of the impact of multimodal features on immersion and cognitive load, with online measures mostly triangulated with offline measures (cf. Doherty & Kruger, 2018 and Kruger et al., 2015 for an overview, and Kruger & Doherty, 2016 for a multimodal methodology to measure cognitive load).

24.5 Offline measurements of the impact of AVT on immersion and cognitive load

Although the individual components of the multimodality of AVT products have received little overt attention, a number of studies have investigated the impact of adding subtitles to film on immersion and cognitive load, and the difference between the effects of subtitles and dubbing on film processing. Perego et al. (2015) provide a strong case showing that subtitles do not have cognitive costs when compared with dubbing and in fact, boost lexical acquisition. They use a range of offline cognitive measures, including a comprehension questionnaire, a dialogue recognition questionnaire, a face-name association test, and a visual scene recognition test. Kruger and Steyn (2014), and Kruger, Doherty, et al. (2017) include offline measures in their studies on the processing of subtitled video. Based on the principles of cognitive load theory, they use the items proposed by Leppink and colleagues (2014, 2015), which identify specific items that can be used in measuring different components of cognitive load. Extraneous cognitive load is of particular relevance for the measurement of the impact of subtitles and other multimodal elements on cognitive load. Extraneous cognitive load relates to the way in which information is presented and is therefore impacted when information is presented in different modes, as is the case with subtitles as well as AD.

Kruger et al. (2016) and Kruger, Soto–Sanfiel, et al. (2017) use a suite of offline measures of immersion (including items on presence, transportation and character identification). Immersion refers to the degree to which an audience becomes immersed, or lost, in a fictional reality. In their study, Kruger, Soto–Sanfiel, et al. (2017) find that immersion (and in particular transportation as the sense of being transported into a fictional reality) increases for a second-language audience watching a television drama with English same-language subtitles. In contrast, however, when comparing dubbing and subtitling, Wissmath et al. (2009) find that dubbing is more
immersive than subtitles. However, they do not compare the dubbed and subtitled versions with a version without dubbing or subtitling, and their audience is accustomed to dubbing rather than subtitling.

The field of AD has produced a number of studies that investigate the interaction of different modes in creating an AVT product that will be accessible to an audience deprived of full access to the visual modes. As mentioned earlier, Wilken and Kruger (2016) investigate the impact of *mise-en-scène* and elements of film perspective on immersion. They find that AD does not always include a description of modes such as perspective provided by camera angles, shot length and editing, which has a negative impact on the ability of a sighted audience (and by extension, a blind or visually impaired audience) to immerse. Fryer and colleagues investigate the related concept of presence in a number of studies dealing with AD. They find that AD enables the blind audience to experience at least as high a sense of presence as a sighted audience, and in some cases an even higher sense of presence (Fryer & Freeman, 2012, 2014; Fryer et al., 2013).

Romero-Fresco and Fryer (2013) and Fryer and Romero Fresco (2014) transfer the theatrical convention (of providing blind audiences with an introduction to the theatre) to film. They investigate the use of audio introductions (10 minutes of description preceding the film that provides information on various elements of the visual modes, such as film style, and descriptions of the context, including characters and setting). The blind audience responds very positively to these introductions.

More recently, these offline measures have been triangulated with more objective online measures of the impact of certain aspects of AVT on the cognitive processing of audiences. This provides a promising avenue for research on the cognitive processing of multimodality in the context of translation.

### 24.6 Online measures of cognition and triangulation

The online measurement of the processing of multimodal texts has developed rapidly over the past decade (see, for example, Holsanova, 2014). The majority of this work in the context of Translation Studies has been done in AVT and in particular on the processing of subtitles by viewers, although a handful of studies also use eye tracking to investigate film processing for audio description using eye tracking. One example is the study by Vilaró et al. (2012) on the impact of sound on the visual processing of film. Another example is the study by Kruger (2012) on the way in which eye movements can provide information on viewer construction of narrative based on multimodal codes.

According to Doherty and Kruger (2018, p. 47), “empirical research on subtitling and captioning has understandably focused on examining their processing and reception by diverse audiences as part of a rich multimodal experience that spans various genres and formats”. Originating in the 1980s in the work by d’Ydewalle and colleagues in Belgium, eye tracking has been used extensively and increasingly to study the visual processing of subtitles in order to get closer to an understanding of the cognitive processing of subtitles together with other visual modes in film. Kruger and Doherty (2018) provide an overview of more than 30 eye-tracking studies conducted in the context of subtitling (the majority of these studies in the past decade). Many of these studies deal primarily with attention distribution between the linguistic mode of subtitles and the visual modes on screen. Perego et al. (2010), however, investigate the cognitive effectiveness of the processing of subtitles by also testing the extent to which viewers manage to process other visual modes. Their study therefore grapples very specifically with cognition in the presence of multimodality.
Fox (2016) engages with another dimension of multimodality by using eye tracking to study the way in which viewers process “subtitles” that were created as an integrated element of the multimodal text. She calls this “integrated titles” and shows how placing the text according to aesthetic principles as well as according to principles of visual saliency (where the eyes would automatically be directed due to the other visual modes) facilitates the overall processing of the film. Caffrey (2012) likewise uses eye tracking to measure how experimental subtitling impact on viewer perception.

Szarkowska et al. (2011) and Szarkowska et al. (2016), as well as Krejtz et al. (2013), make use of eye tracking to investigate the impact of presentation speed, degree of editing and shot changes on subtitle processing, thereby making a contribution to our knowledge of the effects of these multimodal elements on the cognitive processing of subtitles. Although eye tracking provides a valuable measure for investigating cognition in the processing of multimodal translation products, particularly by providing evidence of the visual processing of written language and other visual codes such as images or movement, it has some limitations when it comes to investigating anything more than visual attention distribution. Measures of cognitive load, such as fixation duration and fixation count, are only meaningful when similar modes are compared, such as non-verbal visual modes. Due to the difference in visual processing required when reading text (i.e. many, short, linear fixations) in comparison to scanning a scene or exploring a face for cues about emotion (fewer, longer, distributed fixations), differences in fixation count or duration become meaningless. The nature of film also means that constant changes in the luminosity and location of fixations that alter the shape of the pupil make changes in pupil diameter less reliable than in studies on static stimuli. In this respect, measures such as revisits to different elements of the screen become more important. These limitations mean that triangulation with offline measures is essential and also make it important to explore other online measures.

Electroencephalography (EEG) is one such online measure. It measures electrical brain activity on the scalp to study cognitive processing. In the context of Translation Studies and AVT, EEG offers exciting possibilities for measuring cognitive processing during the reception and production of translation products. However, EEG data has to be triangulated with eye-tracking data and also with offline measures in order to validate what is as yet a largely uncharted methodology.

One early study attempting such triangulation is that by Kruger, Doherty, Fox, & de Lissa (2017). They triangulate alpha power with subjective measures of cognitive load, showing promising results in analysing EEG data as time course data to show how levels of cognitive load change over time. Kruger, Doherty, & Ibrahim (2017) also investigate beta coherence between the prefrontal cortex and the posterior parietal cortex as an online measure of psychological immersion as it fluctuates over the course of a film.

24.7 Concluding remarks

A focus on multimodality and cognition in translation research represents a shift away from the purely linguistic components of translation. This not only changes the discipline fundamentally but has also been necessitated by the proliferation of multimodality in the texts translators have to translate. This is evidenced by the central role online texts play in our lives today, as well as the proliferation of media. This chapter provides a brief (and selective) overview of the way in which multimodality impacts on the production as well as the reception of translation products, and of ways in which to study the impact of multimodality on cognition.

As translation scholars become more versed in the methodologies of disciplines such as psychology, psycholinguistics and cognitive science, the rigour and replicability of research increase.
Translation, multimodality and cognition

opens up exciting possibilities for arriving at a better understanding of the way humans engage with the multimodal reality around them, particularly when this reality involves the complexities of two or more languages in interaction, while building on decades of research on linguistic and cultural aspects. The challenge we face is to disentangle the impact of various modes in varying relationships of redundancy, supplementarity and interdependence on cognition, not only for the translators producing target texts but also for the audiences who have to process the texts.

Further reading


In this chapter, the authors provide an overview of the development of eye tracking in the study of subtitling. They outline some of the main measures used in eye tracking in this context and identify the main uses and limitations of these measures.


This chapter traces the development of the concept of multimodality in Translation Studies. In particular, it engages with the implications for the discipline of viewing translation as a multimodal activity.


This is a foundational text on multimodality, outlining a theory of communication in the context of interactive multimedia. It provides a comprehensive delineation of the boundaries between modes and media of communication.


The authors of this chapter provide a perspective on cognitive Translation Studies that focuses on the reception of translations rather than the production of translation. As such, the emphasis is on the cognitive processing in the minds of readers, listeners and viewers who receive translation, interpreting and audiovisual translation products.


This chapter argues for internal coherence among cognitive approaches to translation and interpreting, and identifies current and future trends in cognitive Translation Studies.


In their introduction to the special issue on multimodality in translation research, the authors engage with the interdisciplinary nature of translation research on multimodality that requires the crossing of various boundaries. In doing so, they highlight the methodological challenges this field encounters.

References


Kruger


Translation, multimodality and cognition


