İSTANBUL UNIVERSITY PRESS
1453

İstanbul Üniversitesi Çeviribilim Dergisi
Istanbul University Journal of Translation Studies

Review Article | Derleme Makalesi

Open Access | Açık Erişim

Multimodal Change in Translation Studies: Transcending the Boundaries

Çeviribilimde Çok Modlu Değişim: Sınırları Aşmak

Check for updates

Sevda Balaman [1]

[1] Sivas Cumhuriyet University, Faculty of Letters, Department of Translation and Interpreting, Sivas, Türkiye

Abstract

Upon recognizing the changing landscape of the communication register in the modern era with the burgeoning impacts of emerging technologies, the scope of translation research has been redefined in a way that embraces the prominence of non-verbal modes, alongside the linguistic code in meaning-construction. In this sense, the multimodality concept has increasingly received attention as an indispensable component of the meaning-transfer process, and its bidirectional connection with *Translation Studies* has been specially emphasized as a new research avenue to capture a clear image of the convergence in these fields. However, there exists a dearth of studies in the related literature, specifically devoted to the functionality of multimodal cohesion in translations. With this in mind, in light of previous research, this review study seeks to determine (a) how the phenomenon of multimodality locates itself in translation, (b) how it structures the extent of translatology research, (c) how multimodal texts are deciphered with reference to well-established transcribing methods, and (d) the current directions that the translation research in this paradigm is moving towards. For all these purposes, starting with an in-depth elaboration of conceptual issues by addressing the underlying core theories, in the methodology section of this paper, a number of theoretical frameworks used in different genres were presented to specify how a multimodal message is best transferred. Overall, this study can add new dimensions to this discipline by enabling translators to gain more awareness of the multimodal stratification of texts and the relevant translation methods.

Öz

Gelişen teknolojilerin artan etkileri ile bağlantılı olarak, modern çağda iletişim dilinin değişen yapısının farkına varılması üzerine çeviri araştırmasının kapsamı, anlamı oluşturmada dilbilimsel modun yanı sıra sözel olmayan biçimlerin de önemini benimseyecek bir şekilde yeniden tanımlanmıştır. Bu anlamda, çok modluluk kavramı, anlam aktarım sürecinin vazgeçilmez bir bileşeni olarak, gün geçtikçe daha da dikkat çekmiştir ve bu kavram ile *Çeviribilim* arasında bulunan çift yönlü bağlantı, söz konusu alanlardaki yakınsamanın net bir görüntüsünü yakalamak için yeni bir araştırma alanı olarak özellikle vurgulanmıştır. Bununla birlikte, ilgili alan yazında, özellikle çevirilerde çok modlu bütünlüğün işlevselliğine yönelik çalışmaların eksikliği söz konusudur. Bu durumu akılda tutarak, bu inceleme çalışması, önceki araştırmaların ışığında, (a) çok modluluk olgusunun çeviride kendisini nasıl konumlandırdığını, (b) çeviri araştırmasının kapsama alanını nasıl yapılandırdığını, (c) çok modlu metinlerin köklü çözümleme yöntemleri referans alınarak nasıl deşifre edildiğini ve (d) bu paradigmadaki çeviri araştırmasının ilerlediği mevcut yönleri belirlemeyi hedeflemektedir. Tüm bu amaçlar doğrultusunda, temel teşkil eden başlıca teorilere değinilerek, kavramsal konuların derinlemesine detaylandırılmasından başlayarak, bu makalenin yöntem bölümünde, çok modlu bir mesajın en iyi şekilde nasıl aktarılacağını belirlemek için farklı türlerde kullanılan bir takım teorik çerçeveler sunulmuştur. Genel olarak bu çalışma, çevirmenlerin metinlerin çok modlu katmanlaşması ve ilgili çeviri yöntemleri hakkında daha fazla farkındalık kazanmalarını sağlayarak bu alana yeni boyutlar kazandırabilir.

Keywords audio-visual translation · mode · multimodality · non-verbal modalities · social semiotics

Corresponding author | Sorumlu Yazar: Sevda Balaman sbalaman@cumhuriyet.edu.tr

Introduction

With the ubiquitous prevalence of digital technologies, which have stepped into every sphere of our lives, communication is no longer limited to the monomodal ways of expression, but embodies a repertoire of multiple modalities including visual, audial or verbal codes (Boria & Tomalin, 2020). In this vein, it seems indispensable to acknowledge the widespread shift from writing's centuries-long supremacy towards the take-over of the other channels in sense-construction alongside language (Kress, 2003). This situation has been explained with the concept of *multimodality*, which broadly stands for the negotiation of meaning through multiple channels (Kress & Van Leeuwen, 2001; Wu & Pan, 2023), expressed in core and sub-modes that coexist in the same instance (Weissbrod & Kohn, 2019). Given that "no text is, strictly speaking, monomodal" (Gambier, 2006, p. 6), because of the existence of visual elements such as layout, cover, or typography (Kaindl, 2013), it is evident that multimodality is inherent to each text (Kaindl, 2020). The multimodal text is constructed by stratifying the sense with various sign systems, each of which has a distinct but complementary nature within the semiotic paradigm (Jewitt, Bezemer, & O'Halloran, 2016). It highlights "the combined communicative potential of multi-semiotic systems that encompass written texts and images, and also different types of technology that incorporate visual, verbal and aural modes" (Li, 2019, p. 10). In line with this, since multimodality is not depicted as "a monolithic concept" (Tuominen, Jiménez Hurtado & Ketola, 2018, p. 4), a multimodal text cannot be confined to one exemplar sort, but covers a wide array of genres ranging from comics, illustrated books, video clips, films, songs (Weissbrod & Kohn, 2019), cartoons (Okyavuz & Sancaktaroğlu Bozkurt, 2022), advertisements (Povoroznyuk, 2022), restaurant menus (Li, 2019), video games (Mejias-Climent, 2017), infographics (O'Halloran, Tan & Wignell, 2016), opera, museum texts (Valdeón, 2024) and so forth.

With the conceptual expansion of the text structure, the scope of various disciplines has been redefined by acknowledging the position of multimodality in decoding and encoding meaning. As an important tool for communication with the target groups, the field of translation is not the exception to this phenomenon, as well (Boria & Tomalin, 2020; Wu & Pan, 2023). In this vein, there has been a repetitive call to reconfigure the mainstream of *Translation Studies* (TS) in the last years (e.g., Boria & Tomalin, 2020; Carreres & Noriega-Sánchez, 2020; Valdeón, 2024), in such a way that regards translation "as the transposition of meaning in the multimodal semiotic landscape of the contemporary social world" (Kress, 2020, p. 27). Kress and Van Leeuwen's (2001, 2006) multimodal communication theory forms the basis of the multimodal turn in the TS discipline, characterized within the "communication-sensitive and culture-sensitive translation" orientations (Kaindl, 2020, p. 54). Within the revisited description of the translation act, the linguistic-only approaches may not thoroughly transfer the meaning in the composite products, embedded in both verbal and non-verbal modalities (Boria & Tomalin, 2020). To this end, the discipline of TS needs to develop suitable strategies, situated in the contemporary multimodal transcribing approaches (Adami & Ramos Pinto, 2020), in order for an effective analysis of multimodal texts where a number of sign systems operate concurrently within the given textual arrangement (Boria & Tomalin, 2020).

However, although this concept has been investigated thoroughly in many disciplines, including literacy studies, linguistics, and semiotics in the recent decades because of the inevitable connection between verbal and non-verbal elements that constitute the core of the new-age communication system (Tuominen et al., 2018; Weissbrod & Kohn, 2019), there is a lack of research specifically devoted to the convergence between multimodality and TS. Therefore, this paper might add to the relevant literature by addressing the

functionality of the phenomenon of multimodality in translation research as an indispensable component of the meaning transfer process.

Aim of the Study

The primary purpose of this paper is to bring to the fore the contemporary turn in TS with reference to multimodality as an emerging area in translation practice. To this end, building on the previous research, this study strives to elaborate issues regarding (a) what functionality the phenomenon of multimodality has in the translation research, (b) how insights obtained from the multimodal theory structure the face of translatology studies, (c) how multimodal textualities are handled for proper meaning transference, and (d) which direction current issues in TS are heading for.

For these aims, this paper starts with describing the theoretical bases of the phenomenon of multimodality and scrutinizes the coverage of multimodal translation based on the relevant underlying assumptions by addressing how multimodality operates in the translation research. Then, this article presents an overview of the transcription frameworks for the discourse analysis of the multimodal outputs. Given that the linguistic-oriented translation methods are not sufficient to achieve the sense consistency between texts, the methodology of this paper is reliant on the presentation of the most common transcribing methods in the literature applied in different genres in order to show how the multimodal stratification of texts can be deciphered in such a way that communicates the complete panorama accorded by various semiotic systems. In this sense, mostly informed by the systemic functional approach (Kress & Van Leeuwen, 2006) and multimodal transcription and text analysis (Taylor, 2003), certain examples from various sorts of genres are presented in order to exemplify the transcription methods. Finally, because the translation is towards machine translation due to the impacts of digitalization, this paper also specifies how multimodality locates itself in machine translation practices by analysing the previous studies in this sense. Overall, this review paper can broaden insights into the multimodal configurations in the translatology research, traditionally prioritizing the verbal modalities over the other non-verbal modes as meaning-carrying sign systems, towards the focus on the multifaceted and multi-semiotic fabric of texts (Boria & Tomalin, 2020; Carreres & Noriega-Sánchez, 2020; Valdeón, 2024).

Theoretical Basis of the Phenomenon of 'Multimodality'

The theoretical strand of the concept of multimodality is based on one particular influential theory, *Multimodal Social Semiotics*, (Boria & Tomalin, 2020, p. 12), which has been proposed in the multimodal discourse analysis, with its grounding on *social semiotic* assumptions and *Visual Grammar* theorization by Kress and Van Leeuwen (2006), originally informed by Halliday's (1978) *Systemic Functional Theory* and *Social Semiotic Theory of Communication*. The Hallidayan approach proposes that meaning resources should be read "within a sociocultural context, in which the culture itself is interpreted in semiotic terms" (Halliday 1978, p. 2) by addressing the "dynamic and context-dependent" notion of signifiers (Pérez González, 2014, p. 127). This theory has two main assumptions. According to the first premise, a society's semiotic resources are the result of their configurations through social (inter-)actions. A particular social semiotic activity can modify the speech acts, images, gestures, and writing forms in different cultures to meet the special needs of both the source and the target society (Kress & Van Leeuwen, 2006). As such, the presumption that "same mode, same meanings" is unsustainable (Kress, 2020, p. 31) due to the diverse ways that modes have been developed in various cultures (Adami & Ramos Pinto, 2020; Kress & Van Leeuwen, 2001). The second assumption contends that any semiotic resource must address meanings that emerge in each of the three distinct but completely interconnected social realms to become a functional tool for communicating in all respects (Kress, 2020, p. 31). These three variables, which are essential for the representation of meaning

within the context, are as follows: *Field* denotes the topic or substance of the discussion, by helping to focus attention on what is happening, when it is happening, and to whom. The *Tenor* indicates the social relationship and the purpose present in a communication exchange between the interlocutors. Finally, *mode* refers to the sense-conveying resource utilized in the representation and communication (Halliday, 1978, p. 62), and the case where more than one mode is simultaneously represented ends up with a multimodal ensemble (Kress, 2010).

In such operations, different modes represent distinct semiotic resources enabling the concurrent manifestation of discourses and types of (inter)action, alongside including varied affordances, with specific potentialities and restrictions that affect how signals are made in representations (Kress & Van Leeuwen, 2001, p. 21). Therefore, it is underlined that each semiotic resource serves for the complete whole in meaning construction (Jewitt et al., 2016, p. 3). Within the tenets of the social semiotics of multimodality, the language modality is just one way of those representations, alongside the other modalities. As such, it no longer enjoys its privileged position as the principal channel for sense-construction and is reduced to just one among several options of encoding and representation of the message (Boria & Tomalin, 2020, p. 14). In short, multimodality has two main premises, one of which requires the coexistence of both verbal and extra-linguistic semiotics within the same context, while the second of which posits that this interaction across modes can only be deciphered from the viewer's standpoint in the cognitive system (Everett, 2015).

An Overview of 'Multimodal Translation'

Historically, TS was firmly tied to the language-oriented paradigm in the 1970s, "which resulted in a merely rudimentary integration of semiotic considerations into the scholarly debate" (Kaindl, 2020, p. 52); however, recognizing the criticality of multimodality (Boria & Tomalin, 2020), the field has experienced a shift "towards culturally sensitive, sociologically motivated, and medially conscious research" (Kaindl, 2020, p. 52). Built on the social semiotics of multimodality (Hodge & Kress, 1988; Van Leeuwen, 2005), translation is no longer seen as the transfer of the message across verbal signifiers, but an act across modes within the multimodal semiotic landscape by shaking the invincible position of language-centeredness (Boria & Tomalin, 2020), traditionally accepted "as its prototypical core area" (Kaindl, 2020, p. 59). Kress introduced the term of *transduction* with reference to the transference of meaning from one semiotic mode to the other one (2010, p. 124). Concurrently, the translator is no longer solely required to entail an expertise in decoding the message through only the linguistic functioning, but also to be competent for simultaneously exploiting the other semiotic affordances that add layers to the sense of the text composition and design in deciphering the meanings (Damaskinidis, 2016; Kaindl, 2020).

The semiotic dimension of the translation has been grounded on the *intersemiotic* translation theory, proposed in Jakobson's (1959) seminal work. Jakobson (1959, p. 233) introduced the tripartite translation taxonomy by signifying the expansion of the translation concept beyond the language focus:

1- Intralingual translation, also known as rewording, refers to interpreting verbal signs in light of other signs within the same language.
2- Interlingual translation denotes interpreting verbal systems across languages.
3- Intersemiotic translation, also titled as transmutation, is about the rendition of verbal signs through nonverbal elements (Jakobson, 1959, p. 233).

Intersemiotic translation is the process of translating the verbal component into the non-verbal sign systems. Then, examples of this would be converting spoken words into pictures, music, or movies. It is important to highlight that intersemiotic translation, in accordance with Jakobson's (1959) original concept, is contingent on the language-focus as the starting point; it does not work the other way around. Therefore,

this conception of intersemiotic translation has been revised by other scholars (e.g., Kaindl, 2013; Toury, 1994) in a way that refers to the translation of two different meaning resources, irrespective of the presence of the verbality, which eventually results in a more comprehensive and wide-ranging treatment of translation "as information mediation", rather than "as language mediation" (Ketola, 2016b, p. 1181). Although Jakobson's (1959) three-part classification has long been criticized because of inconsistencies in the terminological issues and language-centeredness (Kaindl, 2020), it has pioneering impacts on the subsequent research that addresses the semiotic orientation in translation (O'Halloran et al., 2016). Toury (1994) reconfigured intersemiotic translation by incorporating the concept of *intrasemiotic* translation, which addresses the transfer of non-verbal components to the other non-verbal systems, disregarding the dominance of the language-only mode. The remaking of films can be considered in this framework. This issue has then been handled by addressing the phenomenon of multimodality (Kaindl, 2020). Therefore, multimodal translation can be thought as the expanded version of the intersemiotic translation where both the source and/or the target messages are constructed through the juxtaposition of different sign systems (Tuominen et al., 2018), each of which has a sense-conveying potential (Ketola, 2016a), by subtly yet successfully repurposing one another on the same interface (Ketola, 2016b, p. 1188). The multimodal meaning-making mechanism in translation is summarized by Li (2019, p. 11) as follows:

Figure 1

Meaning-making process in multimodal translation (Li, 2019, p. 11, in light of Jakobson's typology, 1959, p. 233)
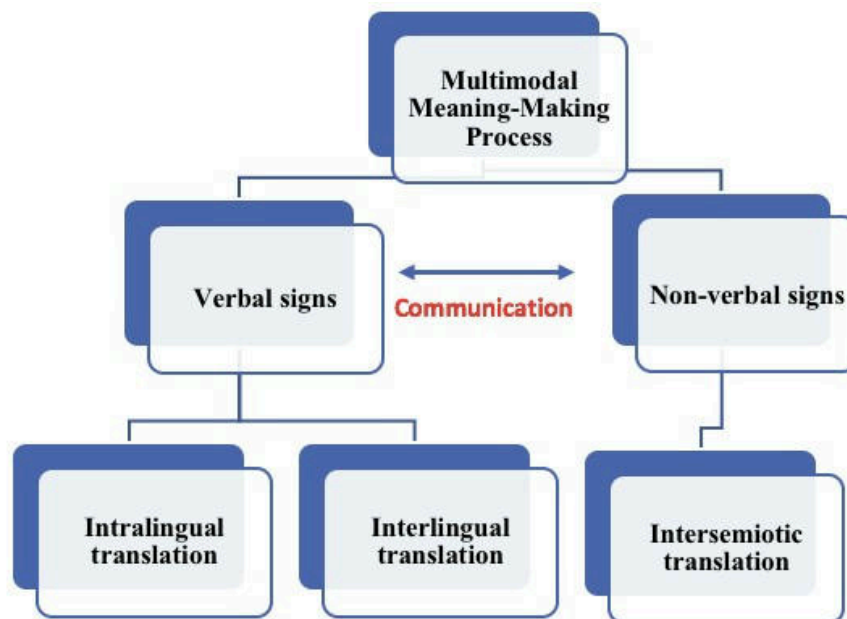
As seen, reconstructing the meaning in a multimodal artefact is reliant on the communication between both verbal and non-verbal elements. In the same token, multimodal translation is not an independent description of modes without considering the cohesion in the products, but a reorganization of signifiers in order to articulate a new thought, by formulating what was previously only hinted at in contemporary translation theory and practice and mostly picked up via experience in a new configuration (Kokkola & Ketola, 2015, p. 227). As a subfield of TS, *Multimodal Translation Studies* addresses the presence and inter-action of various modes in translational situations, which ends up with the analysis of the aligned messages expressed in different modes with each one's complementary competence for the overall sense construction (Tuominen et al., 2018).

Transcription Frameworks for Multimodal Discourse Analysis

In multimodal-based translations, the most outstanding approach regarding the social semiotic dimension of translation is the systemic-functional approach to multimodality (Kress & Van Leeuwen, 2006), also known as "metafunction analysis" (Kokkola & Ketola, 2015, p. 222), which originated from Halliday's *Systemic Functional Linguistics* (SFL) theory, which sees language as it is because of being an agent that functions in the society (1978). The Hallidayan notion posits that language, as a social semiotic mechanism, fulfils the three metafunctions: "ideational, the interpersonal and the textual" (Halliday, 1978, p. 46). The *ideational level* elaborates the way the language represents the external world. The *interpersonal* level deals with the interaction created between the receiver and sender of the message via language. The *textual level* is reliant on the organization of language that results in a cohesive text (Halliday, 1978, pp. 45-46). Because the Hallidayan (1978) perspective centres around the language, it has been broadened by Kress and Van Leeuwen, with reference to other semiotic signifiers alongside language for communicating the message (2001, 2006), by treating each mode "as identical in their meaning-making strategies" (Kokkola & Ketola, 2015, p. 222). This revisited framework has targeted the multimodal analysis of discourses, reconceptualizing three meta-functions as follows (Kress & Van Leeuwen, 2006, pp. 42-43):

Table 1

The metafunctional analysis of multimodal texts (Adapted from Kress & Van Leeuwen, 2006, pp. 42-43)

Meta-functions	
Ideational (Representational) level	based on the elaboration of participants, action, plot, and participants' features to transpose the meaning
Interpersonal (Interactive) level	based on the way the participants of the message interact with themselves
Textual (Compositional) level	based on the intersemiotic mechanism of the elements that carry meaning in the products considering the layout of the text or how the visuals are presented for achieving coherence in meaning

Built on this theoretical foundation, Lu scrutinized the multimodal dimension of a series of films within the specificity of neologisms (2023). For this aim, a checklist was adopted for identifying the content in the three metafunctions (Kress & Van Leeuwen, 2006) listed above (See Table 1). The data revealed that out of the four strategies, i.e., "transliteration, literal translation, explicitation and creation", used for the meaning transference of the neologisms that exist in Chinese subtitles, the explicitation is the most-frequently utilized strategy type (Lu, 2023, p. 11). In this study, Lu concluded that in the subtitled films, both verbal and visual properties complemented and strengthened each other in the meaning composition (Lu, 2023). In addition to this, Chen and Wang (2016) applied an integrated approach for the subtitle translation in which both the SFL-informed multimodal methodology (Kress & Van Leeuwen, 2006) and the semiotic translation framework (Painter, Martin & Unsworth, 2013) were employed for analysing the subtitles in a film. The authors noted the strong interplay between verbal and non-verbal elements in choosing the right strategy for the subtitling procedures. The visual mode not only reinforced the verbal system, but also filled in the cultural gaps between the target and the source language (Chen & Wang, 2016).

Alongside the audio-visual translation (AVT) of films, the SFL-informed framework has also been applied in other genres. For example, Plastina (2022) explored how meaning is reconstructed in visual medical abstracts through "the expression, content and context planes" (p. 32) with reference to Halliday's systemic-functional principles (1978). The research indicated that the selection of specific verbal and visual elements and their semantic links are meant for the provision of mapping between the original core experiential senses and the various meaning-conveying resources represented on the interdependent expression and

content levels. However, it was also revealed that experiential meanings seem to operate as a channel to particularize new interpersonal senses (Plastina, 2022, p. 49).

From a similar standpoint, O'Halloran et al. (2016) also elaborated the poly-semiotic patterns in multi-modal textualities, including infographics, mathematical images and symbols, or videos by conducting a multi-layered SFL-oriented analysis (Halliday, 1978) for delving into the resemiotization stages. The researchers provided a framework based on the analysis of semiotic affordances within the meta-functional systems, as resources of "experiential, logical, interpersonal and textual meanings" (O'Halloran et al., 2016, p. 225). They presented various examples of the way this framework can be utilized by specifying the inter-semiotic dimensions in different instances. Zhao (2023) also investigated the translation process in comics as multimodal texts by delving into the way the illustrations impact the concretization of the target message. The researcher used a framework based on the visual grammar theory of Kress and Van Leeuwen (2006), with reference to Halliday's SFL approach (1978). The results demonstrated that this type of multimodal translation applied to comics is a process of exploitation of visual modes for re-creation of meaning in the text, a justification strategy for choosing certain verbal elements over the others, and giving grounds on the use of specific strategies such as condensation or addition than the others (Zhao, 2023).

Van Meerbergen (2009) also implemented a meta-function analysis for deciphering the translations of picture books by presenting a descriptive model based on the integration of the multimodal theorization by Kress and Van Leeuwen (2006) and translation assumptions by Toury (1995). The analysis revealed that all modes of representation fulfilled the three metafunctions proposed by Halliday (1978). For example, the study noted that the interpersonal interaction in the visuals of picture books, one meta-function of Halliday's (1978) categorization, is created through the gaze vectors between the character depicted in the illustrations and the viewer. Kress and Van Leeuwen's (2006) social semiotic perspectives, which broaden Halliday's (1978) systemic functional notion to the multimodal context, were also adopted in Chen's (2023) study, addressing the multimodal translation analyses of public notices in Macao by drawing data from the self-prepared corpus to indicate the visual-text relationship. The aim of this study was to examine the general tendency of the linguistic and multimodal representation in these products. The data revealed that the visuals in photographs, cartoons or comics accompanied the meaning in the verbal text and diversified the message with strengthened representations. From a very different point of discussion, Ouyang and Fu (2021) also extended the multimodal translation research by investigating the link between the paralinguistic aspects and the meaning-making mechanism in a mock consecutive interpreting practice conducted with one speaker and nine interpreting trainees. The data were collected using an analytical framework, partly situated on SFL assumptions (Halliday, 1978). The results underlined the positive impact of non-verbal paralanguage elements on the quality of the meaning in the outputs. For example, it was noted that having an eye contact with the speaker enabled more accurate and complete renditions.

Another applicability of the social semiotic approach in translation is *the multimodal transcription and text analysis*, which originated from Thibault (2000) and was integrated into TS by Taylor (2003). Taylor (2003) adapted Thibault's (2000) framework to the AVT research, with reference to Kress and Van Leeuwen's (2006) assumptions. This remodelled framework is reliant on the tabular formulation of the multimodal message by segmenting its components into different rows and columns denoting the verbal, visual and audio layers of the output. It entails six columns shown below (Taylor, 2003, p. 196):

Table 2

A framework for the multimodal translation of audio-visual texts by Taylor (2003, p. 196) in light of Thibault's (2000, p. 314) methodology

Time	Visual frame	Depicting Contents of the visual frame	Description of the kinesics action	Soundtrack	Subtitles
Frame-by-frame duration shown in chronological order	Still image of the relevant scene	Detailed portrayal of actions in the visual with respect to camera position, the perspective of the action, the virtual distance of the shot as well as the secondary information about the scene such as visual collocation or colour	Depicting the movements or facial gestures and body actions	Verbal dialogue, music, or sound	Metafunctional interpretation of meaning in different layers of content and expressing them in the accompanying subtitles

The information given in this modelling can be helpful for translators to clarify the particular semiotic modality in which the meaning is conveyed, which eventually enables them to make wiser use of subtitles and reduce the cognitive load on the audience (Taylor, 2016, p. 228). It has four basic principles:

1- determining the most important parts of the semiotic signifiers

2- dividing the text into sections

3- scrutinizing the interaction of modes in and across sections

4- employing a meta-functional analysis (Reviers, 2018, pp. 24-27)

This framework has also been implemented in previous research, mostly in the AVT research (e.g., Reviers, 2018; Taylor, 2016). However, both the SFL-informed multimodal theory (Kress & Van Leeuwen, 2006) and the multimodal transcription theory (Taylor, 2003) prioritize the linguistic mode over the others. According to the metafunction study of images, verbal language and visual representation are structurally similar. Contrarily, multimodal transcription accurately captures the verbalization of each mode, overlooking how those modes merge for conveying meaning (Kokkola & Ketola, 2015). It is also noted that these methods are time-consuming and painstakingly detailed, hence becoming overwhelming for translators. As a response to this case, the phenomenological approach applied to film translation can be highlighted in this sense (Ketola, 2018). With this perspective, the cross-modal impacts are taken into consideration for the deep analysis of the given data in translating the whole. This analysis says that the multimodal genre is more than its fragments where each structure adds dimension and layers to the meaning, hence necessitating the analysis of the modes in a rather atomistic fashion (Kokkola & Ketola, 2015).

Apart from these two approaches, the multimodal dimension in translations was also researched in restaurant menus by Li (2019) based on a corpus-based methodology, formulated upon the synergy between Jakobson's (1959) triadic translation perspectives and Kress and Van Leeuwen's (2006) semiotic notions. In the study, 3000 Chinese dish names and their translations were compared using a corpus, with an emphasis on the interaction between verbal elements and the accompanying visuals. The data indicated a limited application of intersemiotic principles, inconsistent translations between cooking methods and the ingredients, and the untranslated cases of most culture-based dish names. This research suggests the focus on the multimodal aspect of the textuality when translating dish menus for proper cross-culture transfer (Li, 2019).

Ketola (2016b) also scrutinized the multimodal nature of picture books in light of their translations in Finnish based on a framework with reference to the three categorizations by Jakobson (1959). Ketola (2016b) aimed to determine the way the visuals differed from each other in forming the story in different translation excerpts and how the meaning was negotiated through both verbal and non-verbal resources. It was confirmed that the detailed information about the story regarding characters, plot and location was communicated by the choices of illustrations used by the translators. Ketola (2016b) concluded that the meaning transference of the picture books in question is reliant on the interplay of two levels of translation, i.e., interlingual and intersemiotic translation (Jakobson, 1959), but not exclusive of themselves. The text-visual interplay in technical texts was also delved into by Ketola (2016a) in another study via a totally different framework suggested above. The researcher used a "cognitively grounded theoretical framework" (Ketola, 2016a, p. 67), which posits that the reading comprehension stage initiates with the generation of two distinct representations of meaning accorded by the two modes (illustrations and the verbal text). If the conveyed meaning is consistent in the two modes, then making concrete connections between the representations becomes easier, which then results in one compound representation in the cognition. However, if the meaning in the two codes contradicts each other, then building connections will be difficult, which requires more discussion, debate or conciliation between the information yielded by the two modes for proper sense transference. Overall, this suggests that the translation act is inevitably dependent on the information gathered in the negotiation process between the two codes (Ketola, 2016a, pp. 76-77).

Another framework was implemented by Silvester (2018), based on the existing theories for the analysis of subtitled films. Three stages of the holistic contextual analysis are included in the process. It suggests conducting a preliminary analysis of the technical components of the movie and the situation of subtitling. It then moves to breaking down the story, talking to the director, and reading reviews or articles that were published to determine the main ideas and the skopos of the film. Lastly, it offers a careful assessment of the subtitles in context, in connection to the film's skopos, while taking into consideration the main themes of the movie and the cultural setting(s) of its release, based on the material uncovered during these stages of analysis (Silvester, 2018, p. 71).

Finally, Ramos-Pinto (2018) presented a detailed analytical framework for the analysis of the non-standard varieties in the subtitles of audio-visual artefacts, built on a three-strata schema, i.e., "textual, diegetic and sociocultural" (p. 19), which identifies the interdependency of each mode in communicating the meaning in subtitles and suggests that the message hidden in one mode can be transferred via multiple modes. Within this schema, the textual dimension is based on specifying the non-standard varieties and the way they are composed with reference to extra-linguistic dimensions, which impact the selection of neutralization and preservation strategies for decoding the irregular discourse usages. The second layer of meaning, i.e., the dietic dimension, is based on the communicative functionality of the non-standard varieties, in which the multimodality relationship in the target and source texts is analysed through three modes, namely spoken mode (e.g., accent, intonation, or vocabulary), misé-en-scene mode (e.g., costume, setting, behaviours), and subtitle mode (e.g., graphic representations). Finally, the social-cultural aspect concerns the contextual clues that help the reconfiguration of meanings (Ramos-Pinto, 2018).

In sum, as seen from the previous research shown above, multimodal translation is operated by unfolding different semiotic resources that layer the meaning in a range of textual types, with reference to different transcription methods. However, based on this synopsis of prior research, it is evident that the formulation of frameworks for decoding meanings structured via multimodal properties is mostly guided by SFL-informed methodologies (Halliday, 1978; Kress & Van Leeuwen, 2006). It also goes without saying that the multimodal configurations have been specifically handled and emphasized in the AVT research with the

burgeoning interest in audio-visual products (Taylor, 2016). In this sense, Gambier (2023, p. 14) highlights that the AVT research will constitute a prominent position of the multimodal translation in the future (See Figure 2). Overall, it seems that the audio-visual products will increasingly dominate the industry, hence necessitating the systematic analysis procedures particular to that genre for the proper rendition of the input message into the target language.

Figure 2
The status of AVT in multimodal translation research
(SDH= Subtitling for the Deaf and Hard of Hearing)
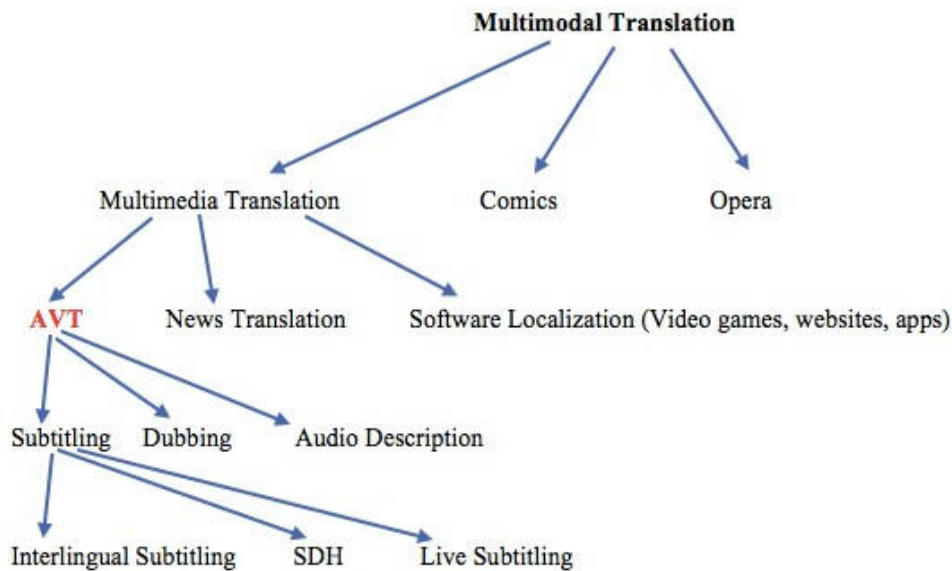
Source: Adapted from Gambier, 2023, p. 14

Current Directions in Multimodal Translation

As suggested above, in order to thrive in this new era, it is required to possess abilities to perceive and generate multimodal texts, and the same applies to the natural language processing (NLP) within the computational landscape in that NLP necessitates the replication of the same competence in order to adopt a human-oriented standpoint and perception in various artificial intelligence (AI) tasks (Nam & Jang, 2024). One such case is the machine translation (MT) that has been extensively used for a long time with outstanding performance, mostly with textual representations. However, given the proliferation of multimodality in everyday communication (Kress, 2010), multimodal machine translation (MMT) has gained momentum, which draws the contextual information from different channels to infer the hidden sense in the texts involving images, videos, captions, or sign languages (Sulubacak, Çağlayan, Grönroos, Rouhe, Elliott, Specia, & Tiedemann, 2020).The main goal of MMT is to leverage the quality of the translation system, addressing the semantic interaction among the auxiliary inputs (Liu, Zhao, Sun, Liu, & Yang, 2021) through "complex input-output mapping" (Sulubacak et al., 2020, p. 30). Especially in cases where structural or semantic ambiguities arise, the aligned semantic link across different modes other than language is helpful for ambiguity resolution and enhancement in the robustness of the system (Çağlayan, 2019).

In MMT, the semantic correspondence is mostly achieved through visual input processing for increased neural MT by fixing ambiguity (Li, Shimizu, Chu, Kurohashi, & Li, 2023; Sulubacak et al., 2020). Therefore, much of the previous research based on MMT has been largely guided by the image-directed MT (e.g., Elliott, Frank, Sima'an, & Specia, 2016; Zhao, Komachi, Kajiwara, & Chu, 2020), extensively relied on image datasets

(e.g., Elliott et al., 2016; Young, Lai, Hodosh, & Hockenmaier, 2014). There also exists other research, tailored to the spoken-language translation (e.g., Akiba, Federico, Kando, Nakaiwa, Paul, & Tsujii, 2004), as well as video-guided MT (e.g., Li et al., 2023; Wang, Wu, Chen, Li, Wang, & Wang, 2019), driven by various datasets in different language pairs (See Li et al., 2023; Sanabria, Çağlayan, Palaskar, Elliott, Barrault, Specia, & Metze, 2018; Wang et al., 2019). Depending on these data sets, different models built in various architectures have been proposed "by first using the global visual features and then moving on to more sophisticated attentive approaches which incorporate spatially aware features" (Çağlayan, 2019, p. 86). The taxonomy of model structures especially for the image input and NLP covers "retrieval-based, templated-based, encoder-decoder-based, and generative-based models" (Nam & Jang, 2024, p. 2).

In short, different software packages and AI-driven systems have been developing day by day for tracking semantic patterns in multimodality. Although many of the tools at present are largely dependent on image captioning models, the scope is evolving rapidly for scrutinizing multimodal texts composed of various modalities in a systematic way as well as developing multimodal literacy. To sum up, it is evident that the MMT constructed in new methodologies has promising outcomes in tackling the burden of multimodality (Nam & Jang, 2024). However, the evaluation issue can be underlined as a drawback in any frameworks of this new-age translating paradigm listed above. While there have been a variety of automatic evaluation computations such as the METEOR, BLEU, and TER metrics, human judgement still stands out as a trustworthy alternative by implementing direct assessment and post-editing strategies to check the elements of fluency, correspondence, and adequacy in translations. For this reason, even if the MMT is considerably evolving in generating appropriate data, it seems not possible to overlook the human-oriented evaluations, which eventually end up with more accurate outputs (Sulubacak et al., 2020).

Conclusion

Overall, depending on the shift in the communication literacy from linguistic orientations to multimodal configurations that result from the proliferation of technological advances, there has been a growing interest in situating the translatology research on the elaboration of how the concept of multimodality manifests within this discipline through well-rooted theoretical principles (Boria & Tomalin, 2020).This indicates a new paradigm that prioritizes non-verbal signifiers, alongside the verbal modalities within the semiotic landscape of translation (Kress, 2020). Therefore, linguistic-only translation methods, traditionally used for the inter-lingual transfer, are not deemed very effective in decoding the message in a multimodal complex, constructed by multi-semiotic layers (Boria & Tomalin, 2020). In this vein, as a critical component of sense transfer, gaining a dose of awareness for the multimodal nature of contemporary texts and the necessity of utilizing systematic transcription methods is essential for reflecting the complete thought in these multi-tier texts, with reference to the auxiliary elements expressed in different signifiers along with language (Kaindl, 2020). Depending on this, the translator competence has been redefined by emphasizing the acquisition of in-depth awareness of multimodality, covering media and genre analysis competency, alongside traditional skills in language and cultural awareness, professional expertise, researching mastery, and sense transferring competence (Kaindl, 2020, p. 65). Additionally, it is highlighted that translators can be in cooperation with other experts in the multimedia industry (Pedro Ricoy, 2012), in line with the demands and specifications brought about by the usage and reception circumstances in the relevant culture (Kaindl, 2020, p. 64).

In sum, to attain the full expertise in this technology-rich contemporary world, translators need to be highly cautious about the functionality of different semiotic signs in adding meaning to the text cohesion and design with their complementary and accompanying roles (Kaindl, 2020, p. 65). Building the heightened sensitivity to the manifold links between linguistic and non-linguistic properties probably ends up with the

selection of more suitable strategies and methods for decoding meanings imparted via different channels by delving deeper into the sense resources (Pan & Liao, 2023). However, because of the rapidly-evolving technologies, new multimodal texts are constantly being added to the multimodal genre repertoire, which calls for new methodological methods for transposing the meaning. Therefore, although this study has detailed certain essential issues regarding the multimodal analysis, which might be helpful for translators to gain a sense of awareness about the criticality of multimodality and has indicated the well-rooted frameworks in the previous literature used for transcribing multimodal discourses, state-of-the-art technologies and new forms of texts necessitate further theoretical and methodological approaches for proper meaning transference. In line with this, future studies can be conducted with the purpose of formulating new analytical tools to be applied in a large corpus of multimodal ensembles for tracking the alignments of multi-semiotic codes and words by installing them on the existing frameworks indicated throughout this paper (e.g., Kress & Van Leeuwen, 2006; Taylor, 2003). Additionally, since the previous research has specifically attempted to bring to the fore the multimodality phenomenon in translation practices, rather than in the interpreting domain with very few exceptions (e.g., Ouyang & Fu, 2021), further research can be carried out for delineating the multimodal stratifications in interpreting input and output formulations.

Author Details	Sevda Balaman (Assist. Prof.)
Yazar Bilgileri	[1] Sivas Cumhuriyet University, Faculty of Letters, Department of Translation and Interpreting, Sivas, Türkiye
	ⓘ 0009-0004-2390-7323 ✉ sbalaman@cumhuriyet.edu.tr

References

Adami, E. & Ramos Pinto, S. (2020). Meaning-(re)making in a world of untranslated signs: towards a research agenda on multimodality, culture, and translation. In M. Boria, Á. Carreres, M. Noriega Sánchez, & M. Tomalin (Eds.), *Translation and multimodality: Beyond words* (pp. 71–93). Routledge: New York.

Akiba, Y., Federico, M., Kando, N., Nakaiwa, H., Paul, M., & Tsujii, J. (2004). Overview of the IWSLT 2004 evaluation campaign. In Proceedings of the 2004 international workshop on spoken language translation (IWSLT), Kyoto, pp. 1–12.

Boria, M., & Tomalin, M. (2020). Introduction. In M. Boria, Á. Carreres, M. Noriega Sánchez, & M. Tomalin (Eds.), *Translation and multimodality: Beyond words* (pp. 1–23). Routledge: New York.

Carreres, Á. & Noriega-Sánchez, M. (2020). Beyond words: concluding remarks. In M. Boria, Á. Carreres, M. Noriega-Sánchez, & M. Tomalin, (Eds), *Translation and Multimodality. Beyond Words* (pp. 198-203), Routledge: New York.

Chen, X. (2023). Representing cityscape through texts and images: translations of multimodal public notices in Macao, *Asia Pacific Translation and Intercultural Studies, 10*(1), 53-70.

Chen, Y. & Wang, W. (2016). Relating visual images to subtitle translation in Finding Nemo: A multi-semiotic interplay. Translation & Interpreting, 8(1), 69-85.

Çağlayan, O. (2019). *Multimodal machine translation. Computation and Language* (Unpublished PhD Dissertation). Le Mans Université, France.

Damaskinidis, G. (2016). The visual aspect of translation training in multimodal texts. *Meta, 61*(2), 299–319. https://doi.org/10.7202/1037761ar

Elliott, D., Frank, S., Sima'an, K., & Specia. L. (2016). Multi30K: Multilingual English- German image descriptions. *In Proceedings of the 5th Workshop on Vision and Language (pp. 70– 74), Berlin, Germany*. Association for Computational Linguistics.

Everett, Y. U. (2015). *Reconfiguring myth and narrative in Contemporary opera*. Bloomington: Indiana University Press.

Gambier, Y. (2006). Multimodality and Audiovisual Translation. Proceedings of the Marie Curie Euroconferences MuTra: Audiovisual translation scenarios (pp. 1-8). Retrieved from http://euroconferences.info/proceedings/2006_Proceedings/2006_Gambier_Yves.pdf

Gambier, Y. (2023). Audiovisual translation and multimodality: What future? *Media and Intercultural Communication: A Multidisciplinary Journal, 1*(1), 1-16. https://doi.org/10.22034/mic.2023.167451

Halliday, M. A. K. (1978). *Language as social semiotic: The social interpretation of language and meaning*. London: Arnold.

Hodge, R. & Kress, G. (1988). *Social semiotics*. Ithaca, NY: Cornell University Press.

Jakobson, R. (1959). On linguistic aspects of translation. In L. Venuti (Ed.), *The Translation Studies Reader* (pp. 232–239). London: Routledge.

Jewitt, C., Bezemer, J., & O'Halloran, K. L. (2016). *Introducing multimodality*. 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.

Kaindl, K. (2020). A theoretical framework for a multimodal conception of translation. In M. Boria, Á. Carreres, M. Noriega-Sánchez, & M. Tomalin, (Eds), *Translation and Multimodality. Beyond Words* (pp. 49-70), Routledge: New York.

Ketola, A. (2016a). Towards a multimodally oriented theory of translation: A cognitive framework for the translation of illustrated technical texts. *Translation Studies, 9*(1), 67–81.

Ketola, A. (2016b). Translating picturebooks: Re-examining interlingual and intersemiotic translation. In Lloyd, P. and Bohemia, E. (eds.), *Future Focused Thinking - DRS International Conference 2016*, 27 - 30 June, Brighton, United Kingdom. https://doi.org/10.21606/drs.2016.225

Ketola, A. (2018). *Word-image interaction in technical translation. Students translating an illustrated text.* Unpublished PhD Dissertation, University of Tampere, Tampere

Kokkola, S., & Ketola, A. (2015). Thinking outside the "Methods Box": New avenues for research in multimodal translation. In D. Rellstab & N. Siponkoski (Eds.), Rajojen dynamiikkaa, Gränsernas dynamik, Borders under negotiation, Grenzen und ihre Dynamik. VAKKI-symposiumi XXXV 12.– 13.2.2015 (pp. 219–228). Vaasa: VAKKI Publications 4.

Kress, G. & Van Leeuwen, T. (2001). *Multimodal discourse: The modes and media of contemporary communication.* London: Arnold.

Kress, G. & Van Leeuwen, T. (2006). *Reading Images: The Grammar of Visual Design (2nd ed.).* Taylor & Francis Group, New York

Kress, G. (2003). *Literacy in the new media age. Modes, logics and affordances*. London: Routledge.

Kress, G. (2010). *Multimodality: A social semiotic approach to contemporary communication*. Routledge.

Kress, G. (2020). Transposing meaning: translation in a multimodal semiotic landscape. In M. Boria, Á. Carreres, M. Noriega-Sánchez, & M. Tomalin, (Eds), *Translation and Multimodality. Beyond Words* (pp. 24-48), Routledge: New York.

Li, S. (2019). A corpus-based multimodal approach to the translation of restaurant menus, *Perspectives, 27*(1), 1-19, DOI: 10.1080/0907676X.2018.1483408

Li, Y., Shimizu, S., Chu, C., Kurohashi, S., & Li, W. (2023). Video-helpful multimodal machine translation. *In Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing (pp. 4281–4299)*, Singapore. Association for Computational Linguistics.

Liu, X., Zhao, J., Sun, S., Liu, H., & Yang, H. (2021). Variational multimodal machine translation with underlying semantic alignment. *Information Fusion, 69*, 73-80. https://doi.org/10.1016/j.inffus.2020.11.011

Lu, S. (2023). How existing literary translation fits into film adaptations: the subtitling of neologisms in Harry Potter from a multimodal perspective. *Visual Communication, 0*(0), https://doi.org/10.1177/14703572221141959

Mejías-Climent, L. (2017). Multimodality and Dubbing in Video Games: A Research Approach. *Linguistica Antverpiensia, New Series: Themes in Translation Studies, 17*, 99–113.

Nam, W. & Jang, B. (2024). A survey on multimodal bidirectional machine learning translation of image and natural language processing. *Expert Systems with Applications, 235*(2024) 121168, 1-14. https://doi.org/10.1016/j.eswa.2023.121168

O'Halloran, K. L., Tan, S. & Wignell, P. (2016). Intersemiotic translation as resemiotisation: A multimodal perspective. *Signata, 7*, 199-229.

Okyayuz, A. Ş., & Sancaktaroğlu Bozkurt, S. (2022). Çok modluluk ve çeviri: Nasreddin Hoca örneği. *Avrasya Uluslararası Araştırmalar Dergisi, 10*(32), 176-193. https://doi.org/10.33692/avrasyad.1176853

Ouyang, Q. & Fu, A. (2021). Effects of non-verbal paralanguage capturing on meaning transfer in consecutive interpreting. In M. Zhang & D. W. Feng (Eds.), *Multimodal approaches to Chinese-English translation and interpreting* (pp. 192–218). Routledge, New York.

Painter, C., Martin, J. R., & Unsworth, L. (2013). *Reading visual narratives.* London: Equinox.

Pan, L. & Liao, S. (2023). Modelling audiovisual translation of non-fiction videos: A multimodal approach to subtitling. In L. Pan, X. Wu, T. Luo, & H. Qian (Eds), *Multimodality in Translation Studies* (pp. 21-34). Routledge: London and New York.

Pedro Ricoy, R. (2012). Multimodality in translation: Steps towards socially useful research. *Journal Multimodal Communication, 1*(2), 181-203. https://doi.org/10.1515/mc-2012-0012

Pérez González, L. (2014). Multimodality in translation and interpreting studies: Theoretical and methodological perspectives. In S. Bermann & C. Porter (Eds.), *A Companion to Translation Studies* (pp. 119–131). Wiley.

Plastina, A. F. (2022). Remaking meaning through intersemiotic translation: the case of online medical journals. *The Journal of Specialised Translation, 38*, 32-54.

Povoroznyuk, R. (2022). Intersemiotics of multimodality: Advertisement in translation. *Journal of Language and Linguistic Studies, 18*(Special Issue 2), 1141-1164.

Ramos Pinto, S. (2018). Film, dialects and subtitles: An analytical framework for the study of non-standard varieties in subtitling. *The Translator 24*(1), 17-34

Reviers, N. (2018). Tracking multimodal cohesion in audio description: Examples from a Dutch audio description corpus. *Linguistica Antverpiensia, New Series – Themes in Translation Studies, 17*, 22–35. https://doi.org/10.52034/lanstts.v17i0.477

Sanabria, R., Çağlayan, O., Palaskar, S., Elliott, D., Barrault, L., Specia, L., & Metze, F. (2018). How2: A large-scale dataset for multimodal language understanding. *In Proceedings of the Work-shop on Visually Grounded Interaction and Language (ViGIL).* NeurIPS.

Silvester, H. (2018). From paratext to polysemiotic network: A holistic approach to the study of subtitled films. *Linguistica Antverpiensia, New Series: Themes in Translation Studies, 14*, 71–83.

Sulubacak, U., Çağlayan, O., Grönroos, S.-A., Rouhe, A., Elliott, D., Specia, L., & Tiedemann, J. (2020). Multimodal machine translation through visuals and speech. *Machine Translation, 34*, 97–147 (2020). https://doi.org/10.1007/s10590-020-09250-0

Taylor, C. (2003). Multimodal transcription in the analysis, translation and subtitling of Italian films. *The Translator, 9*(2), 191–205. https://doi.org/10.1080/13556509.2003.10799153

Taylor, C. (2016). Multimodal corpus analysis of subtitling: The case of non-standard varieties. Target. *International Journal of Translation Studies, 28*(2), 222–236. https://doi.org/10.1075/target.28.2.04tay

Thibault, P. (2000). The multimodal transcription of a television advertisement: theory and practice. In A. Baldry (ed.) *Multimodality and Multimediality in the Distance Learning Age* (pp. 311–385). Campobasso: Palladino Editore.

Toury, G. (1994). A Cultural-Semiotic Perspective. In Thomas A. Sebeok (ed), *Encyclopedic Dictionary of Semiotics, vol. 2 N-Z.*(pp. 1111–1124). Berlin and New York: Mouton de Gruyter.

Toury, G. (1995). *Descriptive Translation Studies and beyond*. Amsterdam and Philadelphia: John Benjamins.

Tuominen, T., Jiménez Hurtado, C., Ketola, A. (2018). Why methods matter: Approaching multimodality in translation research. *Linguistica Antverpiensia, New Series: Themes in Translation Studies, 17*, 1–21.

Valdeón, R. A. (2024). The translation of multimodal texts: challenges and theoretical approaches, *Perspectives, 32*(1), 1-13, DOI: 10.1080/0907676X.2024.2290928

Van Leeuwen,T. (2005). *Introducing Social Semiotics*. London and New York: Routledge.

Van Meerbergen, S. (2009). Dutch picture books in Swedish translation. Towards a model for multimodal analysis. In D. C., Dries (ed.). *Translation and (Trans)formation of Identities. Selected papers of the CETRA Research Seminar in Translation Studies, (pp. 1-20)* Lirias: CETAR.

Wang, X., Wu, J., Chen, J., Li, L., Wang, Y., & Wang, W. Y. (2019). VATEX: a large-scale, high-quality multilingual dataset for video-and-language research. Computing research repository arXiv:1904.03493

Weissbrod, R. & Kohn, A. (2019). Introduction. In R. Weissbrod, & A. Kohn (Eds), *Translating the Visual. A Multimodal Perspective* (pp. 1-12). New York, Routledge.

Wu, X., & Pan, L. (2023). Introduction: Multimodality in Translation Studies: Themes and Models. In L. Pan, X. Wu, T. Luo, & H. Qian (Eds), *Multimodality in Translation Studies* (pp. 1-20). London and New York: Routledge

Young, O., Lai, A., Hodosh, M. & Hockenmaier, J. (2014). From image descriptions to visual denotations: New similarity metrics for semantic inference over event descriptions. *Transactions of the Association for Computational Linguistics, 2*, 67-78.

Zhao, Y., Komachi, M., Kajiwara, T., & Chu, C. (2020). Double attention-based multimodal neural machine translation with semantic image regions. In Proceedings of the 22nd Annual Conference of the European Association for Machine Translation, pp. 105–114, Lisbon, Portugal. European Association for Machine Translation

Zhao, J. (2023). Re-instantiation of multimodal meaning in the translation of Chinese comic Zhuang Zi Speaks: The Music of Nature. *Asia Pacific Translation and Intercultural Studies, 10*(1), 34-52.