Multimodality

Luis Pérez-González

Multimodality, a term first used in the late 1990s (Jewitt et al. 2016:2), is the study of how we make meaning by combining multiple signifying means or modes – for example, image with writing, music and body movement, speech with gesture – into an integrated whole. While more established disciplines such as linguistics, semiotics and musicology engage primarily with those semiotic resources that fall within their respective conceptual remits, multimodality investigates the synergies between co-occurring semiotic resources. From a multimodal standpoint, each mode “is understood as realizing different communicative work” (Jewitt 2014/2017:16) and is moulded by its own semiotic limitations, potentialities and affordances. Multimodal ensembles – whether they adopt the form of a graphic novel, a film, an illustrated textbook, a museum exhibit or an everyday conversational encounter – result from the interplay between the relevant co-operating modes.

Multimodality specialists have “not as yet focused on questions of translation” (Taylor 2016:222), while translation scholars have been slow to engage with multimodal concepts and methods. Theoretical and methodological developments in translation studies until the 1990s were driven by the conceptualization of translation as a process of written language transfer where the printed word is the only signifying means at play. The study of spoken texts as loci of interpreting activity has similarly tended to revolve around their verbal fabric, often glossing over the semiotic contribution of the orality and corporeality of interpreter-mediated speech. This disciplinary emphasis on language-centred meaning-making processes largely derives from the influence of linguistics during the formative period of translation (and interpreting) studies (Baker 2005). It is also consistent with the entrenched prevalence of monomodality, understood as the dominance of one signifying constituent, such as written language, over other types of meaning-making resources.

Approaches to multimodality

Translation studies has not yet managed to articulate clearly how semiotics and multimodality relate to one another. Stecconi (2009:261) argues that, when used in a lax sense, semiotics encompasses “research that goes beyond verbal language”, including “multimedia/multimodal material”, a term he associates with Gottlieb’s (2005) work. But Gottlieb’s (2017:46) own contribution to this debate identifies multimodality, described as an addition to the more established body of scholarship on multi-channel texts and paraverbal translation, exclusively with the work of multimodality specialists (Baldry and Thibault 2006; Kress 2010). Leaving aside the profusion of terms that translation scholars use to designate their theoretical explorations of meaning-making resources and practices, they would appear to agree that multimodality falls within the wider remit of semiotics.
The bulk of the work undertaken to date on the (loosely understood) semiotics of translation aims to draw up inventories of meaning-making signs associated with both verbal and non-verbal sensory channels. Broadly speaking, the structuralist approach to the study of multimodality in translation aims to establish taxonomies in order to “deal systematically with any type of translation encountered in today’s communicative landscape” (Gottlieb 2017:46), whether it concerns text-types such as film where both the original and translated versions are delivered through more than one channel; translations where the target version incorporates new communicative channels that were not used in the original, as in the case of a screen adaptation of a novel; or translations that reduce the number of channels at play in the source, as in a radio adaptation of a film. Jakobson’s intersemiotic translation, defined as the “interpretation of verbal signs by means of signs of nonverbal sign systems” (1959/2000:114) remains a linchpin of these taxonomic explorations, even though scholars have redefined it, to some extent, over the following decades. Remael’s (2001:13-14) use of the term to designate the transfer of meaning across different media, for example, still follows Jakobson closely. By contrast, Fine’s (1984) definition of intersemiotic as the transfer of meaning across medial variants of the same sign – as seen in subtitling, where the meaning of spoken language is conveyed through written text – represents a departure from the original formulation. However, the significance of Jakobson’s contribution to the structuralist strand of multimodal translation research can be best appreciated in Gottlieb’s (2005, 2017) taxonomy of translation types. In this classification, the difference between intersemiotic translations and their intrasemiotic counterparts, the latter understood as textual transfers where “the sign systems used in source and target texts are identical” (2017:51), lies at the heart of a complex web of 34 translation types that encompass “any communicative system working through the combination of sensory signs” (ibid.:50; original emphasis). But while this inclusive theorization of multimodality aspires to account for “any process, or product hereof, in which a text is replaced by another text reflecting, or inspired by, the original entity” (Gottlieb 2017:50; original emphasis), most of the research informed by the structuralist approach has tended to concentrate on translations involving basic transfers of meaning between verbal and non-verbal channels. The collection edited by Poyatos (1997) on various aspects of non-verbal communication and translation, including paralanguage and kinesics, and Chaume’s (2004) structuralist account of signifying means in audiovisual translation are two cases in point. Audiovisual texts, Chaume argues, involve the transfer of meaning along the acoustic and visual channels. The former conveys meaning produced through signs drawn from the verbal code (the spoken word), the paraverbal code (how speech is delivered, rather than what it delivers) and the non-verbal code (including, for example, sound arrangements, special sound effects or music). The visual code, on the other hand, involves the use of signs from the iconographic code (primarily symbols and icons), the photographic code (relating to the use of perspective, colour or light), and the mobility code (operationalized through a classification of proxemic and kinesic cues).

While structuralist approaches deliver comprehensive mappings of signifying resources and allow for thorough and systematic descriptions of how meaning is made and transferred in translation, they also have blind spots. On the whole, they privilege the conventional and static meaning of established signs and codes over the potential of semiotic modes to form and negotiate changing relationships with other modes, as required by the specific demands
of communicative events. This is precisely what the social semiotic approach to multimodality attempts to do.

Among the various strands of multimodal research surveyed by Jewitt et al. (2016:6-13), social semiotics has exerted the greatest influence in translation studies (Kaindl 2012). The central concept in the social semiotic strand of multimodality is mode, defined by Kress (2017:60) as “a socially shaped and culturally given resource for making meaning”. Modes such as “image, writing, layout, music, gesture, moving image, [and] soundtrack” (ibid.) are shaped by the daily social interaction of people, and their semiotic potential is the product of the cultural shaping of the resource in question. A given semiotic resource such as colour may have different meanings in different contexts, as each context “may either have rules or best practices that regulate how specific semiotic resources can be used, or leave the users relatively free in their use of the resource” (van Leeuwen 2005:4). Multimodal texts are therefore composite products resulting from the combination of various modes or modalities that reach the senses of people through media, i.e. “the material resources used in the production of semiotic products and events, including both the tools and the materials used” (Kress and van Leeuwen 2001:22). Examples of media include screens, loudspeakers, paper, fabric, software and clay. Most multimodal texts are also multimedial, as they must be conveyed through various carrier media to be accessed and enjoyed.

While structuralist semiotics foregrounds codes and taxonomies, social semiotics is particularly interested in the social uses and the interrelationships of semiotic modes in social practice. It ultimately seeks to understand how people produce and communicate meaning by combining several modes in a specific social setting. Meaning-making resources can be subjected to scrutiny along three analytical dimensions: discourse, genre and style (Kress and van Leeuwen 2001). Discourses are “knowledges” (van Leeuwen 2005:94) of some aspect of reality and are socially constructed through conventionalized combinations of semiotic resources, or genres, to achieve specific purposes in a specific social occasion. Discourses and genres are highly dynamic constructs, as the social settings and communicative purposes they serve change over time and space. Using distinctive combinations of discourses and genres confers a style or identity on sign users that can be exploited for communicative purposes.

Chueasuai’s (2013) study of the multimodal shifts that arise in the translation from English into Thai of Cosmopolitan’s ‘Love and Lust’ feature articles is one of the earliest applications of social semiotics to translation studies. By incorporating a range of changes in typography, colour, layout, photographic images and the actual text, Cosmopolitan’s Thai edition delivers a euphemized version of this sexually explicit material. Chueasuai focuses on the cultural shaping of the modes at play to account for these changes. The editorial policy of Cosmopolitan’s Thai edition, local media legislation and socio-cultural norms on femininity are shown to influence the discourses, genres and styles articulated in the translated material. The extent to which social settings influence styles, and hence the semiotic potential of the modes used to make meaning, is also explored in Chang’s (2015) study of note-taking in interpreter-training contexts. While traditional studies of note-taking conventions emphasize the fixed meaning that trainers have arbitrarily allocated to a standard set of signs, the social semiotic approach explains how the interpreters’ personal communicative intentions drive their choice and use of resources drawn from a range of visual modes. Significantly, the fact that interpreters tend to exploit the semiotic potential of their chosen modes in a consistent
manner suggests that notes can be conceptualized as a “third visual language with its own logic and meaning-making practices” (ibid.:7).

The social semiotic approach is also at the heart of work that seeks to understand how inter-modal connections are established within multimodal ensembles, in order to dissect holistic perceptions of multimodal texts as unified semiotic entities. One example is Remael and Reviers (2018), who focus on the study of filmic multimodal cohesion, defined as “any instance of implicit or explicit ‘sense-relation’ between two or more signs, from the same or different modes, within a given text that helps the reader create a coherent textual semantic unit” (ibid.:260). Maintaining or (re)creating multimodal cohesion is particularly important in audio descriptions and subtitles for the hard of hearing, as these forms of translation can only deliver fully monomodal versions (acoustic and visual, respectively) of a multimodal original. Remael and Reviers’ study foregrounds the importance of genre, as conceptualized within social semiotics, for the perception of multimodal cohesion, given that viewers’ identification, tracking and understanding of the basic narratological building blocks is driven by past and conventionalized spectatorial experiences. The cognitive dimension of multimodal cohesion has also been explored from a social semiotic standpoint by Ketola (2016), who focuses on the translation of illustrated technical texts and aims to gauge the impact of the degree of alignment between images and written texts – i.e. the extent to which they complement or reinforce each other – on translational decisions. She concludes that “different types of images affect the translation of verbal text in different ways” and that “readers themselves decide, whether consciously or unconsciously, to what extent they process images” (ibid.:77, 78). The insight that the semiotic space created by the written word and the illustrations between and around them varies in each case is consistent with the social semioticians’ dynamic conceptualization of signifying resources, whose meaning-making potential is realized differently in each communicative event.

The most comprehensive application of social semiotic theory in translation studies is Pérez-González (2014b), who draws on Stöckl’s (2004) operationalization of multimodal theory as a networked system of choices. Here, audiovisual texts are seen as produced by mobilizing those visual and auditory modes that will best realize the creator’s communicative intentions. Among the range of visual and auditory modes identified by multimodal theory, image, language, sound and music are regarded as core modes, as each has more than one medial variant – respectively: static/dynamic images; speech, animated and still writing; sound effects and spectrograms; performed music and score/sheet music. The overarching networked system of modal choices is hierarchically organized, so that each core mode commands a set of associated sub-modes. Deploying a specific mode therefore opens up a new system of sub-modal choices. For example, using speech, one of the media variants of the language mode, prompts more delicate selections in terms of volume, intonation, accent, voice quality, rhythm, speech and pauses that filmmakers can use to enhance dramatic effect or characterization. Drawing on a range of case studies, Pérez-González (ibid.) illustrates the extent to which an enhanced awareness of the multimodal distribution of meaning in audiovisual texts can be crucial in informing translational decisions.
Research themes

One strand of research has focused on layout and typography. O’Sullivan (2013) draws on Flood’s (1993) study of Martin Luther’s translation of the Bible to demonstrate how the Wittenberg editions published during the Reformation used roman and gothic typefaces in strategic ways, with the former serving to associate “certain [negative] Biblical elements with the Church of Rome” and the latter being the preferred option to deliver “positively connoted words” from a Reformist perspective (O’Sullivan 2013:5-6). In the context of subtitling, layout and typography have been shown to enable new forms of interaction between the diegetic world and audiences in silent films and mainstream television broadcasts (O’Sullivan and Cornu 2018; O’Hagan and Sasamoto 2016); facilitate deaf viewers’ monitoring of interactional structures and turn-taking changes in films and television content (Neves 2005); foster a sense of shared affectivity among fansubbing networks and their audiences (Pérez-González 2007a, 2014b); and articulate new and creative forms of media content production (Fox 2016; Romero-Fresco 2018).

A second strand of research examines multimedial spaces such as theatres, cinemas and museums, where a number of carrier media are used to stage multimodal texts. In these contexts, meaning-making activities may adopt different intersemiotic configurations, not all of which entail interlingual transfer. Managing and optimizing the use of multimodal resources is crucial to evaluating the performability of theatre plays, which demand “a dramaturgical capacity to work in several dimensions at once, incorporating visual, gestural, aural and linguistic signifiers” into the (translated) staged version (Hale and Upton 2000:2). Pérez-González (2007b:13-14) offers a multimodal analysis of the Spanish dubbed version of 12 Angry Men (1957) in which he illustrates how camera perspective and the focal length of lenses can be exploited to enhance dramatic tension at crucial junctures of the plot and to evoke emotional responses that influence the translation of the dialogue. Similarly, Maszerowska has drawn attention to the contribution that lighting makes to “the saturation of the audiences’ imaginations, complementing and carrying on the plot, reflecting the characters’ points of view and, at the same time, filling in the gaps between dialogues” (2012:83). The semiotic implications of local spatial constraints, object-text interdependence, overall aesthetic coherence, co-presence of material and virtual artefacts, as well as intertextual relations across individual exhibits are some of the issues that have attracted the attention of translation scholars who theorize museums as multimodal ensembles (Sell 2015; Neather 2012, 2017; Liao 2018). From an audience perspective, subtitling for the hard of hearing, respeaking and audio description have increasingly become foci of applied multimodal research seeking to accelerate the integration of individuals with sensory impairments in mainstream theatre, film and museum audiences (Neves 2018; Romero-Fresco 2011; Jiménez Hurtado and Soler Gallego 2015).

The iconic-verbal link in printed texts is one of the most established areas of multimodal research in translation studies. Kress and van Leeuwen’s (1996) grammar of visual design has supported a range of studies on the translation of advertising material (Munday 2004; Millán-Varela 2004; Torresi 2004) and magazines (Chueasuai 2013), all of which conceptualize the partnership between verbal and visual structures as a means to identify particular interpretations of experience and forms of social interaction. The impact of the semiotic dominance of images has also been studied in relation to comics (Zanettin 2008), where
humour is often encoded in the images, rather than the speech balloons. The possibility of altering or removing images when translating comics (Zanettin 2011) can be exploited, for example, to reconfigure the semiotic realization of humour across modes and to explore new forms of interaction between images, colour and written text (Kaindl 2004). The multimodal analysis of static images in comics can also yield a better understanding of the semiotic potential of less studied modes, including spatial orientation, body posture, gaze and distance (Borodo 2015). In illustrated literature, including graphic novels, images play a crucial role in restricting or amplifying certain aspects of the meaning conveyed by the written text (Alvstad 2008), to the extent of being constitutive of plot (Leighton and Surridge 2008). The modification or removal of illustrations during the translation of a book have been found to disrupt the text-image connection observed in the original, and affect the readers’ experience of the books in question (Oittinen et al. 2017).

A further strand of research has demonstrated that the affordances of digital technology allow for experimental approaches to the translation of polyvocal texts that combine multiple layers of verbal and nonverbal material. Translations of such texts may adopt the form of interactive artefacts such as DVDs which combine written texts, photographs, sound recordings and ethnographic field material in ways that facilitate meaningful reading experiences through hyperlinked connections (Milsom 2008). In her attempt to translate French transgender memoirs into English, which lacks the grammatical resources available in the original language to articulate fluid, sometimes indeterminate, expressions of sex and gender identity, Rose (n.d.) exploits hypertext to display multiple translation options for key markers of transgender discourses, thus fostering a palimpsestuous pluralization of sexual and aesthetic experiences and signifying “the queerness of all identity and text” (ibid.). This “technologically-mediated sense of translation” and the intersemiotic reading experiences that it provides have also been explored by Lee (2013:241) in the context of art installations and cyber-poetry. The emergence of further technological affordances in the future, Lee argues, will pave the way for new forms of creative transposition whose study will have to be informed by a “multimodal perspective on translation and a translational perspective on multimodal expression” (ibid.:254).

Alongside printed and digital multimodal ensembles, embodied multimodality has become an important research theme. Studies on para-verbal, prosodic and kinesic behaviour, such as Poyatos (1997), offer insights into the semiotics of the human body, understood as “the use of para-verbal signs – including but not limited to voice qualities, cadence, inflection, or rate of speech – and non-verbal signifiers – such as gestures or movements” (Pérez-González 2014a:122/123). Bosseaux (2015) demonstrates that the interplay between the qualities of characters’ voices (timbre, pitch) and their physical appearance often makes an important contribution to dramatic characterization, as is also the case with a range of sociolinguistic traits embodied in speech and sound, including accent and other markers of linguistic variation. Staging a translated theatre play or dubbing a film often involves decisions that may alter aspects of characterization and erase the sociolinguistic resonances of voice as a marker of identity, which may prove detrimental to the reception of the translated play or film (Queen 2004; Mingant 2010; Bosseaux 2018). In triadic medical encounters, the simultaneous use of various modes can help doctors to elicit patients’ informed consent (Bührig 2004), claim/retain the interactional floor and negotiate conflicting expectations. Understanding these sites of embodied multimodality, where spoken language is complemented by facial
expressions, gestures or body positioning, is crucial for the development of effective interpreting strategies through a “co-ordinated manipulation of several semiotic resources” (Pasquandrea 2011:477). Analyses of embodied multimodal resources such as gaze (Davitti 2015), gesture, body position, proxemics and object manipulation have been conducted in pedagogical settings (Davitti 2012), and suggest that dialogue interpreting actively contributes to constructing “different participation frameworks throughout the encounter” within a specific ecology of action, understood as “the relationships between the participants and the surrounding environment” (Davitti and Paquandrea 2017:105). Similar studies have focused on interpreting in business meetings (Bao-Rozée 2016) and group work dialogue among deaf and hearing students (Slettebakk Berge 2018).

**Future directions**

Multimodality is still in its infancy and needs to build consensus around key aspects of its core theoretical framework. The feasibility of subsuming all signifying means under a finite range of modes and the formalization of the processes through which individual modes become integrated in a single unified ensemble remain subject to intense debate. Cognitive and neurolinguistic models of translation (Pérez-González 2014b:102-110) are developing the means to understand how translational decisions are influenced by the cognitive salience that individual modes gain against the multimodal whole. For instance, the combined use of a single set of visual resources with variations of the same soundtrack will influence viewers’ gazing trajectories when they try to process the multimodal ensemble (ibid.:105). In addition to multimodal transcriptions (Taylor 2016) and multimodal corpora (Soffritti 2018), eye-tracking is yielding the most effective insights into the reception of multimodal texts. Studies of users’ visual behaviour reveal how they process and comprehend a multimodal text and should, in due course, help translators understand how best to construct a similar coherent whole (Kruger 2012). As it continues to grapple with fast-paced media developments and capitalize on wide-ranging methodological advances, multimodality is delivering increasingly robust accounts of the generative power of semiotic resources and offering valuable insight into ever more complex textualities.

**Further reading**


**References**


