1. Introduction

Multimodal artefacts, including audiovisual texts, result from the interplay between different types of co-operating signifying resources that have to be “described and describable together” (Kress and Ogborn 1998, quoted in Iedema 2003: 39), not least when accounting for the transformations such texts undergo through translation. But while terms like ‘multimodality’ or ‘multimodal theory’ feature ever more frequently in translation studies research, the interface between both disciplinary domains remains woefully under-explored. As Taylor (2016: 222) notes, multimodality theorists have “not as yet focused on questions of translation”, while translation scholars have not yet begun to make use in a systematic and sustained fashion of the concepts and methods that multimodality has to offer — despite the widely held perception that the latter provides a robust inter- and transdisciplinary platform to support research across the humanities. As far as audiovisual translation research is concerned, this divide has been exacerbated by the entrenched prevalence of ‘monomodality’ — understood as the dominance of one signifying constituent, typically written language, over other types of meaning-making resources or modes used in a text — in what have been traditionally perceived as the most prestigious forms of cultural or artistic expression and “[t]he specialised theoretical and critical disciplines which developed to speak of these arts” (Kress and van Leeuwen 2001: 1). Until relatively recently, audiovisual texts and their translations have not featured among such artistic manifestations or in the discourses that the latter mobilise.
A robust challenge to the dominance of monomodality, however, has been mounted by the scale and depth of recent changes to the medial environment in which translations are produced and received. Over the last three decades, technological advances triggered by the shift from the print to the screen, and then to digital culture have been instrumental for the emergence of new textualities where written language routinely interacts with still/moving images and/or sound in new and productive ways — whether within physical artefacts such as magazines or textbooks, or in the context of hypertextual assemblages. The impact of these transformations in “materiality and its cognates, mediality and technicity” (Littau 2016a: 82) on the semiotic make-up of digital texts has not gone unnoticed by translation scholars. Their heightened awareness of the materiality of communication — encompassing “all those phenomena and conditions that contribute to the production of meaning, without being meaning themselves” (Gumbrecht 2004: 8, quoted in Littau 2016a: 83) — has fostered deeper engagement with the dialectic between media and semiotic resources and placed the material history of translation under renewed scrutiny (Littau 2016b).

In this context, calls for “an integrated approach to translation” that gives careful consideration to ‘the physical support (stone, papyrus, CD-ROM), the means of transmission (manuscript, printing, digital communication)” and “how translations are carried through societies over time by particular groups” (Cronin 2003: 29) are congruous with efforts undertaken in other disciplinary domains to develop conceptual and methodological tools for the study of communicative practices in all their semiotic complexity (Iedema 2003).

Against the backdrop of growing engagement with the study of multimodal texts, this chapter sets out to chart the changing ontological contribution that subtitles have made to the multimodal fabric of audiovisual texts produced since the silent film era to our days. A number of multimodal issues associated with subtitling have been reported on in previous publications that do not frame themselves explicitly as multimodal (see Pérez-González 2020 for a detailed survey), and a growing body of subtitling scholarship that subscribes explicitly to multimodal theory (Remael 2001, Taylor 2003, Gambier 2006, Mubenga 2009, Pérez-González 2007, Desilla 2012, O’Sullivan 2013, Taylor 2014, Taylor 2006, Gambier and Ramos Pinto 2016) has explored the disciplinary connections between multimodality and audiovisual translation in some more depth. This chapter offers a novel approach to the study of the multimodal dimension of subtitling practices, examining the organic enmeshment of subtitles within the overall semiotics of audiovisual content across successive media cultures, moulded by evolving configurations of technology, power, ontological status and social practices. The chapter aims to yield an in-depth understanding of the implications that the shift from the dominant narrational regime of Western modernity to postmodern aesthetics emerging in the context of digital culture has for multimodal textualities and their translation. This study is therefore not based on the analysis of small-scale translation
shifts in a multimodal context. Instead, it engages with wider theoretical and programmatic debates in translation, multimodal and cultural studies to inform future reflective practice at this disciplinary interface.

The chapter is structured as a critique of the changing role of subtitling in the media landscape, charting the gradual emancipation of this modality of translation from the dictates of the film industry as it comes to serve the more democratic forms of spectatorial media ‘prosumption’ associated with digital culture. Kress and van Leeuwen’s (2006) distinction between ‘recording’ and ‘synthesising technologies’ provides the conceptual network required to trace the shift from an ‘ontology of referentiality’, based on the hegemonic narrational regime, to the ‘ontology of deconstruction’ that lies at the heart of the digital media ecology, shaped by variable degrees of convergence between industrial and amateur practices. Developments in the digital arena, as instantiated by Chinese danmu, are presented as contemporary instances of ‘representation-as-design’ driven by rhetorical choices (Kress and van Leeuwen, ibid.) and analysed in terms of the affordances of ‘semiotic software’ (Djonov and van Leeuwen 2017) and the social semiotic practices that the latter enables. Ultimately, this study shows how, where ordinary rhetors have gained greater visibility and agency, their performance of citizenship involves the deconstruction of representation by exposing the cultural and social make-up of multimodal ensembles.

2. Subtitling in the Era of Recording Technologies

As is also the case with users of other multimodal texts, mainstream film audiences must make sense of meaning distributed across various semiotics. The range of signifying means deployed by film creators to encode multiple strands of acoustic and visual meaning effectively constitute “a single unified gestalt in perception” (Stöckl 2004: 16). Presented with these unified semiotic entities, film viewers have to make connections across different types of signifying resources in a routinised, subconscious manner – drawing on their accrued spectatorial experience to decode conventionalised ensembles of sign types (ibid.). The demands originally placed on viewers have grown as film semiotics began broadening its repertoire of meaning-making resources as early as in the pre-sound film era. The relentless drive to heighthen the appeal of motion pictures as commodities led silent film creators to attempt and convey mute diegetic speech to their audiences in various ways – including, but not limited to, film editing techniques and the use of live narrators, ‘speaking actors’, title cards and intertitles (Dwyer 2005, Pérez-González 2014a, O’Sullivan and Cornu 2018).

The advent of sound in the late 1920s further cemented the status of film as the first modern audiovisual form of entertainment and paved the way for the development of
subtitling and dubbing in a bid to preserve and increase the commercial flow of films across linguacultures. Subtitling is widely held to disrupt the multimodal configuration of the original artefact, as conceived by filmmakers, for the ex post-facto incorporation of snippets of text to a finished film calls for additional inter-modal connections that viewers of the original ensemble were not required to draw. While commercial subtitles delivering translations of the original dialogue do not generally influence either the “pacing and rhythm” of the film or its “intellectual and emotional content” (MacDougall 1998: 168), the need to process written language superimposed on the image represents an additional cognitive effort that may detract from the enjoyment of audiences in dubbing-dominated markets. The impact of dubbing conventions on the co-deployment of semiotic resources in films will not be pursued further in this chapter due to space constraints.

Although the partnership between film and subtitles spans approximately one century, we are only beginning to gain reliable insights into the implications that the superimposition of snippets of written text on images has for the processing of subtitled films. The burgeoning body of eye-tracking research on the behaviour of viewers’ gaze shows that the on-screen mobilisation of subtitles always draws attention from viewers, even when audience members are able to understand the dialogue they are presented with (d’Ydewalle and De Bruycker 2007). After reading the subtitles (Jensema et al. 2000), gaze trajectories generally alternate between the subtitles and the images they are inscribed on, as viewers adjust their processing strategies according to the genre they are watching (Perego et al. 2010). Most viewers spend more time reading the subtitles than looking at the images, although the fixations – understood as the pauses between eye movements, during which the viewer’s eyes remain static over an area of interest – on the latter are normally longer (Perego et al. 2010, Kruger 2018). Effectively, once viewers have read the subtitles in full, they tend to focus their attention on key areas of interest – primarily faces, when these are present – for as long as possible. The evidence emerging from eye-tracking research thus confirms that viewers of subtitled films have less time to explore visual semiotic resources, although audiences in subtitling-dominated audiovisual markets are able to watch subtitled programmes relatively effortlessly – as they are regularly exposed to this specific multimodal configuration (d’Ydewalle and De Bruycker 2007).

Crucially, this incipient body of evidence was not yet available during the 1930s and 1940s, at the time when the industry was developing the set of subtitling practices that, as acknowledged by film scholars, have remained virtually unchanged to date (James 2001). The conventions in question seek to achieve full synchrony between the duration of successive blocks of diegetic speech and the on-screen display of the subtitle(s) conveying their meaning into the target language – assuming standard (albeit adjustable) sets of reading conditions associated with different viewer profiles.
Under these ‘spatio-temporal constraints’ imposed by the film medium, spoken dialogue must be substantially condensed to fit into compact, tightly worded subtitles, which have been shown to have a detrimental impact on the viewers’ perception of characterisation, humour, irony and other culture-specific aspects of subtitled film (Mason 2001, Remael 2003, Desilla 2012). Drawing on Stöckl’s (2004) conceptualisation of multimodal resources, to be elaborated on below, the process of semiotic transfer known as industrial subtitling therefore revolves almost exclusively around two ‘medial variants’ (written and spoken) of the same ‘mode’ (language). Except for those instances where the decoding of non-verbal semiotic resources requires culture-specific knowledge (see featured example 6.1 in Pérez-González 2014a: 188-190), it is generally assumed that viewers will manage to activate any connections across different types of semiotics that the subtitled film requires, just as the original film audience members did.

Critical theory and film studies scholars with an interest in the genealogy of subtitling practices, however, have taken the view that industrial subtitling practices are not, strictly speaking, the necessary product of spatio-temporal constraints associated with the film medium. Instead, the cult of synchrony underpinning those subtitling conventions is argued to be consistent with other semiotic developments in the production of motion pictures, not least the evolution of early presentational films — i.e. “the films of the cinema of attractions” where “theatrical display dominates over narrative absorption” (Grochowski 2003: 56) — into representational ones — which seek to create a self-contained diegetic world driven by narrative motivation allowing viewers to suspend their disbelief. In addition to the deployment of sophisticated editing and montage practices, the construction of this self-enclosed fictional world is heavily contingent on the use of synchronous diegetic sound because it provides an allegedly “unmediated access to [the diegetic] reality” (Minh-ha 2005: 129). Synchronous diegetic sound allows producers to shift viewers’ attention away from the tools and relations of production — i.e. the spaces between image, sound and text that can be seen in the films of the cinema of attractions — and minimises the potential for subjective spectatorial experiences. So while sound synchrony would, on the face of it, aim to promote an aesthetic of objectivity, the centrality of the diegetic in modern film production ultimately signals the industry’s interest in spreading hegemonic commercial discourses that work through linear narratives to control film reception.

Drawing on Gunning (1986) and Musser (2006), Grochowski explores the financial and political drivers behind the expansion of the Hollywood studios, gauging the extent of their impact on the consolidation of the narrational regime. For Grochowski, the synchrony-driven switch from the presentational to the narrational was the outcome of a “struggle between producers of motion content and exhibitors (where the production of narrative films becomes merely a strategy, a ‘tool’ to use by the producer...
Ultimately, sound synchrony, and the shift from the presentational to the narrational paradigm that it brought about, enabled the US film industry to impose a taste for homogenising, easier-to-export linear narratives, as part of a wider synchronicity-led initiative to dominate the global film market – often to the detriment of incipient narrative conventions emerging elsewhere in the world, that were soon replaced by their imported American counterparts (see McDonald 2006 and Chung 2007 for accounts of how this acculturation process played out in Japan and Korea, respectively). Iranian film and cultural studies scholar Hamid Naficy (2003: 193) has also examined the dynamics of film narrative colonisation and revealed the geopolitical import of this industrial strategy. In Naficy’s words, at the heart of the US policy of technological transfer and development aid for the Third World since the 1950s, was this notion of homogenisation and synchronicity of the world within Western consumerist ideology. This is a shift from the earlier policy of diachronicity, promoted by colonists, which tended to keep the developed and the under-developed worlds apart. The emerging form of post-industrial capitalism sought synchronicity in the interest of creating global markets.

Hollywood’s imposition of the dominant narrational regime of Western modernity as a means to exercise maximum control over the reception and commercial performance of films required the enforcement of hierarchical power structures between authorship and viewership, and the technologies available at the time played an instrumental role in facilitating the expansion of this hierarchical regime of signification. The mutually influencing relationship between the technological and the social is aptly articulated by Heath, who notes that “[c]inema does not exist in the technological and then become this or that practice in the social; its history is a history of the technological and social together . . ., in which the ideological is there from the start” (1981: 6). Unsurprisingly, the period during which the blueprint of American film narratives was drawn up featured the rise and consolidation of ‘recording technologies’, one of the three historically successive types of production technologies identified by Kress and van Leeuwen (2006). Aimed primarily at the eye and the ear, recording technologies “allow more or less automated analogical representation” of what was presented as unmediated reality (ibid: 217), thus providing the ideal inscription tools to enable the shift from diachronicity to synchronicity that underpins the narrational regime. Recording technologies placing synchrony at the centre of the temporal/spatial order favour homogenising narratives and uphold the authority of the creator within existing media power structures. Acting as a form of ‘suture’ (Heath 1981), they foster an ontology of representation restricting the range of relations that a film is constructed to establish with its viewers, ultimately endorsing an aesthetic of objectivity based on referential expressions of reality (Silverman 1988). Indeed, the emphasis of cinematography on synchronous diegetic sound as a way to keep the machinery and production process behind filmmaking hidden is typical of the “ontologies of referentiality, a view of representation being founded on direct, referential relations between the representations and the world” (Kress and van Leeuwen 2006: 218).
In order to facilitate the implementation of this industrial/commercial strategy, the centrality of synchrony had to be extended to the subtitling of US film commodities. Minh-ha (1992) coins the term ‘suture subtitling’ to foreground the impact that the generalization of the narrational film regime had on subtitling and the processes of multimodal meaning transfer that it entails. The interplay between power structures and recording technologies during the formative years of subtitling allowed the film industry to “collapse, in subtitling, the activities of reading, hearing, and seeing into one single activity, as if they were all the same”, and hence “to naturalize a dominant, hierarchically unified worldview” (ibid., quoted in Nornes 1999: 18). Under the aesthetic of objectivity associated with recording technologies, subtitles have contributed to upholding the ontology of referentiality, acting as gatekeepers between the diegetic and the extra-diegetic. The role of subtitles in the era of recording technologies has been conventionally confined to delivering in the target language approximate linguistic representations of the verbal meanings and or communicative intentions encoded in the source text. The emphasis has been thus placed on the original message that filmmakers intend to convey through their characters’ speech — the assumption being that translators should privilege the ‘primordial’ meaning expressed in the film dialogue. Further, while the contribution of visual and acoustic semiotic resources to the overall meaning of audiovisual texts is ostensibly acknowledged by audiovisual translation scholars (Desilla 2012, Taylor 2003, Pérez-González 2007, 2014a, 2014b, 2020), professional subtitlers working for the film industry during the era of recording technologies have seen the scope of their mediation, and hence the exercise of their professional discretion, reduced to one single form of semiotic transfer involving the re-encoding of speech through written language.

3. Subtitling during the Period of Ontological Transition

Although the hegemony of the narrational regime continues to place stringent constraints on the multimodal configuration of commercial films, a growing body of documentary and ethnographic cinematic practices is revealing the extent to which “the sound track, like the image track, involves mediations, translations, and representational practices that push the sound into the realm of ideology” (Naficy 2004: 141) – thus framing suturing synchrony as a commercial choice, rather than the only option available under the medial constraints of film. The formal organisation of Minh-ha Trinh’s Surname Viêt Given Name Nam (1989), for example, “question[s] the representational nature of film” by breaking down mainstream filmmaking techniques and pointing to their flaws, therefore preventing the audience from suspending their disbelief (Vietnamese Cinema Blog 2015). Just as Trinh’s filming style tampers with the
norms of cinematic realism (for example, by using lighting to disrupt the mood of a scene, camera movements that wander away from the subject and sound overlapping techniques), it also draws on

superimposed titles and subtitles extensively, graphically and critically. Their large numbers and varied contents and layout give this film a truly calligraphic accent. Throughout, subtitles consisting of the translation of the film’s dialogue and voice-over and of Vietnamese poetry and proverbs are displayed, as is customary, in the lower third of the screen. However, on many occasions, what the diegetic women say in Vietnamese or in heavily accented English is superimposed in different layouts, as blocks of English text on various regions of the film frame, including over the characters’ faces. These graphic titles, or what Trinh calls ‘visualized speech’, act as traditional subtitles by aiding spectator comprehension . . ., [although] they also serve other graphic, critical, and deconstructive functions (Naficy 2004: 145-46).

The ontology of referentiality is also critiqued in ‘accented films’ shot by members of exilic and diasporic communities since the 1960s (Naficy, ibid.). Drawing on artisanal production techniques, accented creators adopt a performative stance involving subjective film shots (ibid.: 136) and other techniques that enact a free indirect style in their films. Significantly, the deployment of counter-hegemonic filmic discourses based on sound misalignment and de-centred compositional patterns – that allows accented filmmakers to “comment upon cinema and reality, instead of just recording, reporting on, or representing reality” (ibid.: 146) – calls for the use of profuse on-screen titles to channel “the problematic of linguistic, cinematic, and exilic translation and displacement” (ibid.).

While the erosion of the boundary between the diegetic and the extra-diegetic can be accelerated in documentary and accented films by displaying subtitles deliberately out of synchrony, among other multimodal strategies, commercial films may also seek to undermine the narrational regime by exploring new forms of interplay between the story and the audience. Romero-Fresco (2018), for example, explores new subtitling approaches where, in contrast to what happens in conventional films and drama series, the material dimension and the function of on-screen titles are conceptualised prior to the production of the multimodal text they will be embedded in. Creative subtitling, as defined by Romero Fresco (ibid.), adopts and extends certain features of ‘ethnographic subtitling’ (MacDougall 1998), primarily its capacity to foster new narrative mechanisms through which the authorial voice bypasses the diegetic characters to engage directly with the audience. Creative subtitles — variously referred to in the literature as ‘decotitles’ (Kofoed 2011), ‘authorial titles’ (Pérez-González 2012), ‘impact
captions’ (Sasamoto 2014, O’Hagan and Sasamoto 2016) or ‘integrated titles’ (Fox 2016) — do not seek to deliver translations of the original spoken dialogue, but to enhance aspects of characterisation or progress the narrative in more involving or immersive ways (Dwyer 2015).

Peter Kosminsky’s 4-part drama The State (2017) – which dramatizes the experiences of four British men and women who leave their lives behind to join Islamic State in 2015 – illustrates the innovative use of authorial titles in mainstream media content. As the new recruits arrive in Syria and settle into the organisation, men begin combat training and the women are introduced to the society’s strict rules under the guidance of commanders and house leaders, respectively. English-speaking characters going about their tasks make use of Arabic sentences pertaining to various rituals and cultural conventions associated with Islam, as a way to signal and reinforce their shared collective identity and community affiliation. When such expressions are used, stylised titles – that look “too aesthetic to be merely informative” (Crawley 2017) – display the transliterated Arabic sentences and their English translation in varying areas of the frame. For example, when Shakira Boothe – a British doctor and single mother who joins the Islamic State in the hope of working in a state hospital – makes arrangements to see her first patients, she is told she cannot work without a ‘mahram’, at which point the title “mahram | male guardian” appears on screen. Another example of commercial film subverting the hierarchical power structures between the narrative and on-screen titles is Paolo Sorrentino’s Il Divo (2008), a biographical drama about Giulio Andreotti, one of Italy’s most controversial Italian politicians who found himself at the centre of a conspiratorial shady network that eventually led to his prosecution. Although the film sets out to re-create the reality of key historical events, various semiotic resources are deployed from the performative perspective of auteurship. This is particularly striking in the opening sequence that flashes a series of murder sequences in quick succession, complete with stylised captions providing details of the victims and the circumstances surrounding their death. Clarke (2009), for example, draws attention to the “artful captions that float behind buildings and emerge from car doors – offering indigestibly comprehensive descriptions of their [the victims’] role in the debasement of Italian politics”, while Jenkins (2009) notes the resemblance between the playfulness of Sorrentino’s captions and those used in Danny Boyle’s Slumdog Millionaire (2008). As Marcus (2010: 253) elaborates,

[t]he pretense to investigative seriousness is immediately undercut by the playfulness of Sorrentino’s captions – the upside-down hanging corpse of Roberto Calvi [one of the murder victims] is accompanied by a correspondingly upside-down caption, which turns right-sideup as the photogram rotates 180 degrees to its proper position. Similarly, the caption identifying Michele Sindona, poisoned in his jail cell, is shown
backwards until the camera circles around the dying man so that we can read his name from left to right. Throughout the film, captions dance about the screen in a kind of freefloating semiotic abandon. Words, like images, have come untethered from their ‘signifieds’. By transforming the convention of the caption – hitherto understood as an indicator of its film’s unproblematic and unequivocal referentiality – into one more sign to be manipulated at will, Sorrentino asserts his ambiguous relationship to the truth claims of the documentary genre, and its offshoot in the films of cinema politico (emphasis in the original).

The authorial titles deployed in The State and Il Divo enable the transgression of traditional narrative boundaries through ‘narrative metalepsis’, as conceptualised by Genette (1983 [1980]) – see O’Sullivan (2011) and Pérez-González (2014a) for a more extended analysis of how subtitling enables this form of metalepsis. Once they are incorporated into the mise en scène, authorial titles become an additional narrative plane intersecting with the unfolding diegetic events – dramatising the blurring of the boundary between fiction and reality. The creators of The State and Il Divo use authorial titles to transgress traditional narrative conventions and exploit the ‘double-layeredness’ of filmic communication (Vanoye 1985) to intrude in the diegetic action. The vertical level of interpersonal communication (between the film-makers and the audience) is prioritised in places over its horizontal counterpart (between the diegetic characters), as titles provide viewers with information and knowledge about the fictional world that would normally be confined to the diegetic characters. The combined use of cinematography and metaleptic captions in Il Divo, for instance, disrupts classical hierarchies in film narrative, undermining the ontological status of fictional subjects. As Hogan (2009) argues, criminal allegations are not made against specific individuals through diegetic speech. Instead, they are communicated to the viewer through the interplay between visual signifying means and performative titles, some of which “are spat out of the mouth of a victim of poisoning, or swing from under Blackfriars Bridge along with the cooling corpse of [Italian banker] Roberto Calvi” (ibid.).

The previous examples have shown how the deployment of titles can be used to destabilise the ontology of referentiality for various purposes, even in the context of media content produced for commercial release. Rather than simply contributing to the representation of a linear narrative or “hiding behind the pretense of an unacknowledged spectator” (Gunning 1993: 5), authorial titles comment upon the diegetic and extra-diegetic realities, promoting the same ‘parataxic’ reading experience as the primitive intertitles featuring in films of the cinema of attractions (Grillo and Kawin 1981). Whether it is through the delivery of diegetic subtitles out of synchrony or in unusual regions of the frame (as is the case with Minh-ha’s ‘visualised speech’), or through the deployment of authorial titles that challenge the privileged status of the
diegetic narrative over its viewers (creating new spectatorial experiences where creators interpellate the audience directly), films and serialised dramas are presenting audiences with new processing demands. Instead of coherent single gestalts, viewers are faced with multimodal artefacts where the strands of meaning conveyed through the juxtaposition (rather than integration) of different signifying means must be dissected and interpreted independently – thus heightening the potential for subjective interpretations of the diegetic narrative.

4. Subtitling in the Era of Synthesising Technologies

The advent of digitisation has brought about the emancipation of subtitling from the media industries. In the digital culture, media content is increasingly shaped by variable degrees of convergence between industrial and amateur practices, with growing numbers of ordinary people becoming involved in its appropriation, annotation (including subtitling), editing and distribution (Pérez-González 2014a, 2017; Dwyer 2017a). Driven by the ubiquity of relevant technological tools, amateur (often participatory) subtitling has quickly reconfigured individual viewership as part of wider processes of social interaction around the production and consumption of media content, turning the “spectatorial culture into participatory culture” (Jenkins 2006: 41). Under this new postmodernist aesthetic, where subtitled media content is often watched via social media platforms, “net users [including amateur subtitlers] are increasingly disinterested in the metanarrative (content) and instead become obsessed with multiple small narratives (comments)” (Xu 2016: 439). In this context, media content represents an opportunity for the negotiation of intersubjectivity – understood as a shared and reciprocal perception of the meaning created through joint activity and socially managed interaction – among members of virtual, geographically dispersed communities of interest. In selecting what is to be subtitled and steering decision-making processes on the materiality and wording of their subtitles, these agencies subordinate the authority of the original to the subtitles’ capacity to mobilise political or playful affinity through the online comment facility that accompanies the video. Participation, in short, takes over spectatorship.

Amateur subtitling is typically executed by virtual communities that capitalize on the affordances of networked communication to exploit their members’ skills sets or collective intelligence. Subtitlers involved in this form of ‘co-creative labour’ (Banks and Deuze 2009) draw on their linguistic proficiency – which does not necessarily equate with translation competence – and activist orientation. ‘Aesthetic activism’ (Pérez-González 2014a), one of the two main strands covered in the literature, is best illustrated by ‘fansubbing’ — a term traditionally associated with the prolific global subculture built around the subtitling of Japanese anime by fans of this genre and, in
more recent times, with a form of immaterial subtitling labour undertaken by followers of commercial dramas, series or musical genres (Barra 2009). Fansubbing practices are normally fuelled by a desire to tamper with commercially translated media content that fans-turned-translators regard as “culturally odourless” (Iwabuchi 2002: 27). Fansubbing practices have been shown to experiment with the formal dimension of subtitles (Dwyer 2012) and bestow a high degree of visibility on translators — who maximise the deconstructive affordances of digital technologies to act as intercultural brokers between the text producers and users from inside the frame. On the other hand, the archival, annotation (through subtitles) and recirculation of media content are also central to the activities of networks of politically engaged amateur subtitlers, for digital culture has facilitated the rise and consolidation of networks of activist subtitlers seeking “to elaborate and practise a moral order in tune with their own narratives of the world” (Baker 2006: 481). Pérez-González (2010, 2013, 2016) has shown how participatory networks of activist subtitling resist the dynamics of the news media industry and challenge the control that global corporations have traditionally exerted over the distribution and consumption of the content they produce. These participatory communities take on the role of self-appointed translation commissioners and selectively appropriate the content they wish to subtitle and redistribute among strategically targeted linguistic constituencies that otherwise would not have been able to access that content through mainstream or commercial media circuits. In the same vein, Baker (2018) has explored how subtitlers embedded in various activist collectives contribute to exploring prefigurative principles (including horizontality, solidarity and rejection of representational modes of practice), choosing to bring their mediation into light, “rather than obscure it and lull the viewer into the illusion of a direct, unmediated experience” of the subtitled content (ibid. 460-461).

As was also the case with aesthetic activists, politically engaged subtitlers ultimately work with their viewers towards the co-construction of affinity spaces for the negotiation of their individual and collective identities. Unconstrained by commercial and financial interests, amateur subtitling by ordinary citizens is contributing to foreground a postmodern aesthetic based on “the primacy of our mundane day-to-day experiences over that of the rigid structures” favoured by the media industry and the narratives that embody their cultural and economic hegemony (Xu 2016: 439). This change in the role that subtitling plays within the multimodal configuration of audiovisual texts is consistent with other manifestation of ‘citizen media’ practices, which encompass

[t]he physical artefacts, digital content, practices, performative interventions and discursive formations of affective sociality produced by unaffiliated citizens as they act in public space(s) to effect aesthetic or socio-political change or express personal desires and aspirations, without the involvement of a third party
or benefactor. It also comprises the sets of values and agendas that influence and drive the practices and discourses through which individuals and collectivities position themselves within and in relation to society and participate in the creation of diverse publics (Baker and Blaagaard 2016: 16).

As noted above, citizen subtitlers – a term adopted here to encompass ordinary citizens-turned-subtitlers working with different agendas – subscribe to a postmodern aesthetic of mundane playfulness or political engagement that often ignores the constraints of narrative motivation and the concomitant need for synchronous temporal and spatial relationships between the original speech and its subtitled version. Indeed, the status of the original is subverted by the comment culture that amateur subtitling practices seek to foster and stimulate. Positing spectatorial subjectivity through parataxic reading practices is at odds with the ontology of referentiality underpinning the production of classical films and therefore requires the use of ‘synthesizing technologies’ (Kress and van Leeuwen 2006). Against the crisis of representation that the shift from recording to synthesising technologies has prompted, digitisation enables alternative forms of ‘representation-as-design’ (ibid.) that open up new possibilities for more democratic spectatorial engagement by bringing the world of the story closer to the space of the audience. The ontology of referentiality associated with the narrational regime is thus superseded by an ontology of deconstruction under which ordinary citizens can watch and produce subtitles, and the diegetic world is no longer closed off to viewers – who can make use of digital technologies to experiment with and expose the multimodal circuitry of the multimodal text, “deconstructing the combinatorial possibilities and laying bare their cultural/social sources” (ibid.: 219).

5. The Multimodal Implications of Design and Rhetoric in Chinese Danmu

In digital media habitats, where amateur rhetors have gained greater visibility and influence, subtitling is emerging as a meaning-making practice that contributes to the deconstruction of the original text by exposing the interstices around its multimodal wiring. As reported in the literature surveyed in the previous section, the ever closer alignment between amateur subtitling and other manifestations of citizen media entails the adoption of experimental practices that mobilise a range of signifying means to style on-screen text in unconventional ways for maximum effect; in terms of the ontological value that they attribute to the source text, citizen subtitlers steer us away from the individual or subject position of their professional counterparts, towards collective discursive spaces of translatorship involving complex negotiations of mundane or ethical identity.
Technological developments – whether in the form of proprietary web-based captioning tools such as Amara, that have been made available to support volunteer initiatives as a way of accruing symbolic capital in the community; freeware editing and subtitling applications that citizen subtitling groups have chosen to integrate within their collective workflows; or social media platforms providing the tools to develop a participatory comment culture around the consumption of subtitled output – have been instrumental in delivering this ontological shift. By enabling users to “select from a range of different semiotic resources” as well as “incorporat[ing] and represent[ing] knowledge about what constitutes effective use of these resources” in citizen subtitling, these tools qualify as ‘semiotic software’ (Djonov and van Leeuwen 2017: 567). As befits synthesising technologies, semiotic software has its own semiotic regime, insofar as the “rules for using the semiotic resources available within a software product are externalised in and tacitly imposed by its design” (ibid.: 571). Significantly, semiotic software is not only important for enabling multimodal affordances and setting restrictions in terms of how various resources – including, but not limited to, colour, font, texture, or animation – can be jointly deployed (van Leeuwen et al. 2013), but also for being semiotic artefacts in their own right with the capacity to influence social practices (Poulsen et al. 2018: 596). The dialectic between the affordances and constraints imposed by semiotic software and the social practices that it enables will be explored in the remainder of this chapter through the lens of Kress and van Leeuwen’s (2006) well-established analytical notions of ‘design’ and ‘rhetoric’. In the context of this study, the former designates the multimodal meaning-making potential of the technological resources facilitating the work of ordinary people involved in participatory subtitling, whereas the latter will be used to foreground the scope and complexity of the social relations that obtain around the production and consumption of subtitled media content.

As the practices associated with successive fansubbing turns (Dwyer 2018) and political subtitling (Pérez-González 2010, 2013, 2016; Baker 2018) are relatively well documented in the literature, the remainder of this paper examines the interplay between semiotic technology (design) and social practices (rhetoric) – and the ontology of deconstruction that they bring about – as instantiated in the context of the under-explored Chinese participatory subculture known as danmu. Howard (2012) describes danmu – also known as ‘barrage commenting’ (Li 2017 and Dwyer 2017b) – as a video annotation system that was first launched in 2007 by a Japanese ACG (anime, comic, game) video-sharing website. The system, which was introduced in China in 2010 by digital platforms like Bilibili.com, involves the display of viewer-generated titles on the very screen where the media content in question is being played. The titles – which, as elaborated below, can deliver translations of the diegetic dialogue or, alternatively, viewers’ comments on a range of issues more or less directly connected with the video they are inscribed on – are therefore superimposed on the multimodal text, rather than
in a separate section, as is the case in YouTube. According to Androutsopoulos (2013), *danmu* — a term that designates both the platform allowing for the generation and sharing of comments, as well as the comments themselves — constitutes a form of ‘participatory spectacle’ where media content is reduced to a simple background, with the production and reading of comments taking over as the focal attraction. Indeed, the climatic parts of a series episode or film may bring about so many *danmu* that comments may form a multi-layered ‘bullet curtain’ (which is how *danmaku*, the original term used in Japanese to refer this practice, translates into English) that may become thick enough to block the visuals (see Figure 1).

![Figure 1. Screenshot of video material featuring *danmu* (Bilibili.com).](image)

In what follows, the analysis of the *danmu* aesthetic as a form of representation-as-design will be informed primarily by Stöckl’s (2004) typology of multimodal resources — a framework that, whilst not primarily developed to analyse audiovisual texts, provides a productive and systematic set of analytical tools based on Stöckl’s conceptualisation of semiotic resources as constituents of hierarchical networks of choices.

As is also the case with mainstream and citizen subtitling, the *danmu* titling culture mobilises a range of modes, i.e. “sign systems from which communicators can pick their signs to realise their communicative intentions” (Stöckl 2004: 11). Language and image remain as core modes – understood as those types of signifying means that are “deeply entrenched in people’s popular perceptions of codes and communication” (ibid.: 14) – in the *danmu* environment, although the prominence of the user-generated titles
signals very clearly (i) how traditional relationships between authorship and viewerwhip are being subverted; and (ii) the extent to which “the inner layer (metanarrative) loses primacy to the surface layer (nonnarrative)” (Xu 2016: 448) that viewers generate. The fact that *danmu* titles seek primarily to negotiate a perceived sense of intersubjectivity among viewers, rather than to enjoy the media content per se entails a clear asymmetry between the volume of information conveyed through each of the two sensory channels at play in this site, with the written word (i.e. the medial realisation of the language mode within the visual channel) taking priority over its auditory counterpart. The inscription of *danmu* on top of the audiovisual text foregrounds the crisis of the representational regime and, more specifically, the emancipation of titles from the primordial narrative.

As is also the case in other forms of citizen subtitling, the Bilibili platform – i.e. the semiotic software driving the *danmu* culture – allows for a range of multimodal enhancements through the activation of one or more title submodes, namely “the building blocks of a mode’s grammar” (Stöckl 2004: 14). Font and colour choices, for example, are among the most productive submodes that viewers can use to style written text and to construct a visual identity for themselves amid *danmu* that have been previously posted by other viewers. Interestingly, in this environment characterised by the proliferation of superimposed comments flowing horizontally over the video, from the right to the left of the screen, it is static written text that becomes relatively salient and commands a higher degree of ontological authority. Static *danmu,* which can only be mobilised by Bilibili premium users – feature at the top and bottom of the frame (Wu et al. 2018) to deliver explanations on both general and specific aspects of the video (Ma and Cao 2017) and to deliver translations of the diegetic speech by way of traditional subtitles (Wu et al. 2018), respectively.

Capitalising on the affordances enabled by semiotic software, members of the *danmu* subculture act as rhetors making cumulative decisions pertaining to the deployment of semiotic resources among the network of submodal options available at any given point. The *danmu* conglomerate therefore emerges as a composite artefact moulded by the decisions made by users during the design stage – as is also the case with other forms of citizen subtitling. Ma and Cao, for example, report that users playing specific roles (e.g. uploader) can stick to the same colour “when commenting repeatedly throughout the video” (2017: 779); groups of posters involved in a shared task, such as translating the lyrics of a diegetic song into different languages and dialects through *danmu* may choose to exhibit a shared visual identity through the adoption of the same colour-cum-font pattern (ibid.); and individual users wishing to interact as part of a self-contained group can signal their shared membership by prefacing their *danmu* with strings of words. In these cases, then, the selection of fonts and colours may seek to reflect what is being talked about, or the relationship between the interactants, just as
in other forms of citizen subtitling. But the deconstructing role of synthesising technologies in the danmu platforms also has some liberating effects: submodal choices are not bound by the need to ensure optimum visibility and readability conditions, as in chaotic barrage commenting environments “the consumption of the metanarrative [the diegetic plot] is cast away in favour of the irrelevant and carnivalesque aspect” of users’ danmu.

From a temporal perspective, there is a difference between static, subtitle-like danmu featuring at the bottom of the frame – which are delivered in synchrony with diegetic speech – and non-translational danmu occupying the entire screen and often blocking out the actual video. Users wishing to post danmu do so by linking the display time for their contribution to the specific time code of the video moment that it comments on. New danmu can be added all the time, but previously logged ones will be replayed every time the video is watched, thus turning the interface into an ever-growing archive of user-generated contributions. As Li (2017: 248) explains, “the temporality structured by the danmaku interface leads to a collective user experience of ‘virtual liveness’ or ‘quasi-liveness’: while all past danmu are shown simultaneously, even though there might be important gaps between the times in which they were inscribed, it is impossible for users to engage in live conversations. Danmu posted in response to a previous comment may never be read by the latter’s author, so interacting this this environment effectively amounts to debating “with someone who spoke in the past but whose words always appear in the present” (ibid.).

In overstepping the boundaries between the diegetic and the extra-diegetic, danmu exhibits many similarities to other forms of title-based mediation driven by a deconstructive ontology examined in previous sections – including authorial titles, understood as a transitional regime between recording and synthesising technologies, and citizen subtitling, as performed by communities galvanised around both fandom or political affinity. But while the network of social relations that emerge among participants in citizen subtitling has been theorised in some depth (Pérez-González 2014a), the rhetorical dimension of danmu has not been explored to date. In the remainder of this section, the scope of the discussion on the rhetorical construction of danmu will be limited to a specific dimension, namely the politics of ‘mutual recognition’ (Thrift 2009).

Amateur subtitling communities have been found to resemble other participatory networks that operate in the citizen media culture insofar as they index “self-referential properties” to their preferred “values, beliefs and practices” (Deuze 2006: 71). Enhancing the visibility of amateur prosumers, in their capacity as representatives of a collective identity shared by subtitlers and their audiences, is therefore crucial to the proliferation of participatory subtitling. Essentially, the drive to enforce an aesthetic or
a politics of mutual recognition between subtitlers and viewers empowers amateur translator networks to adopt an interventionist mediation approach, moulding media experiences to fit viewers’ expectations. By drawing on the affordances of synthesising technologies and reflexively engaging in the manipulation of media content that circulates in their environment, ordinary people are able to establish new participatory sites for the expression of subjective spectatorial experiences and promote, as is also the case with other citizens engaged in forms of self-mediation that do not involve translation, “new forms of playful citizenship, critical discourse and cosmopolitan solidarity” (Chouliaraki 2010: 227).

Danmu take the politics of recognition much further, as illustrated by the abundant textual traces of viewer investment in and engagement with the video playing in the background. Viewers have been shown to rely on danmu for different purposes (Wu et al. 2018). In some cases, they post comments serving the same function as traditional subtitles, i.e. delivering Chinese versions of the original diegetic speech – that, as noted above, will often end up featuring near the bottom of the frame. The collaborative nature of the subtitling process undertaken through danmu has been documented by Wu et al. (ibid.), who report on groups of users collaboratively creating Chinese ‘subtitles’ for different sections of the same video by inscribing their danmu on the relevant fragments. Plotters, on the other hand, make use of danmu to issue warnings about imminent plot twists, typically by framing such climatic events in terms of the emotional reactions they are likely to elicit. Examples include: “‘High-energy reaction ahead!!!’ (i.e., the following plot is very exciting), ‘High-energy reaction ahead, please wear a helmet!!!’ (i.e., the following plot is very exciting and intense, please be prepared mentally), [and] ‘Tut tut, see the end from here’ (i.e., the end can be guessed through this plot)” (ibid.: 212). Enforcing community regulation norms and participating in ‘parasocial interaction’ – where users engage in conversation with imaginary interlocutors or with diegetic characters – are other purposes that danmu may serve. The great bulk of danmu inscribed on any video, however, contributes to the materialisation of a site that enacts the aesthetic of mutual recognition, where viewers become visible to each other when attempting to learn foreign languages, sharing jokes or criticisms, and other more or less mundane experiences. As a textually and formally disruptive contrivance moulded by the interplay between design and rhetoric, danmu environments contribute to the formation of spaces of expressivity and interventionist mediation enabled by the co-existence of diegetic (translational) titles and extra-diegetic posts. In their search for mutual recognition, viewers of media content in the digital build communities of rhetors by exploring the affordances of the semiotic software and gauging the potential and limitations of barrage commenting.
6. Conclusion

As the ontology of referentiality is replaced by one of deconstruction, synthesising technologies are exposing the crisis of the ‘representational’ role that subtitles have played within traditional audiovisual ensembles. Although experimental subtitling practices can also be observed in mainstream media content, the proliferation of citizen subtitling practices and participatory environments like *danmu* is foregrounding interventionist reconfigurations of multimodal textualities driven by aesthetic and/or political allegiances. Further, the dynamics of subtitling in the digital culture are also drawing attention to the extent to which semiotic software – a semiotic artefact in its own right – is influencing the practices through which viewers choose to engage with texts and with fellow viewers.

The popularity of *danmu* platforms raises important questions for subtitling studies scholars regarding the interplay between *danmu* that serve the same function as a translational subtitle and other more parasocial comments, barrage titling, and the obscuring of the multimodal text that they are inscribed on; problematises the long-standing contention among industry players that subtitles represent a form of intrusion in the audiovisual ensemble. Debates concerning the need for commercial subtitles to prioritise clarity and legibility – not least by jettisoning verbal references to acoustic and visual signifying resources that mainstream audiences can access directly – should gauge whether, and to what extent, viewers’ capacity to establish inter-modal connections are being affected. Even if we accept that *danmu* are ushering in “a form of postmodern aesthetic that favours eclectic, user-driven experiences that are detached from the actual meaning or narrative of the products they are actually consuming” (Xu 2016: 447), understanding the specific impact of calligraphic overflow and spectatorial saturation, and (ii) examining how a potential contagion of the barrage culture outside China might affect the production of the primordial audiovisual narratives are bound to emerge as important disciplinary concerns.

The development of these new textualities which, as noted above, enable the collaborative production and critique of subtitles may open new avenues for the expansion of translation crowdsourcing models that have not been exploited to date. Equally important are the methodological challenges involved in the study of community formation by tracing and connecting multiple overlaid layers of written text whose authors and identities are not yet retrievable. Beyond these specific issues, the emergence and popularisation of *danmu*, so far restricted to China, will provide translation scholars interested in the internationalisation of disciplinary discourses with an opportunity to theorise this new regime of signification based on an “aesthetic of hyperflatness” (Xu 2016: 447) and informed by East Asian artistic traditions that works not only towards “the direct subversion of the metanarrative but the very nascency of the Chinese postmodern consciousness” (ibid: 449).

References


Filmography


