

INTRODUCTION¹

Bridging the Gap between Pragmatics and Multimodality

AOIFE BEVILLE¹, FABIO CIAMBELLA², JONATHAN CULPEPER³
¹UNIVERSITY OF FOGGIA, ²SAPIENZA UNIVERSITY OF ROME, ³LANCASTER UNIVERSITY

1. Integrating two disciplines

Traditional pragmatics, rooted in the works of such scholars as J.L. Austin (1962), John Searle (1969), or H. Paul Grice (1975), among others, provides a strong framework for understanding how speakers use language in context to achieve social goals. Developed during an era of ‘page-based’ and face-to-face linguistic dominance, these foundational theories largely overlooked multimodal analysis. Austin was a notable exception; as early as the 1950s, he pointed to the role of “gestures” and “non-verbal actions” in utterances’ meaning-making (1962, p. 76)² – effectively gesturing toward multimodality without naming it as such.³ Nevertheless, traditional pragmatics stopped short of developing explicit multimodal frameworks. It acknowledged other modes in theory but failed to systematise multimodal approaches to the study of language in context.

It is a widespread scholarly view that communication has become intensely multimodal, driven by new technologies such as Artificial Intelligence (AI). This shift is so significant that, as noted by experts in the Visual Communication Project (ViCom), it is paving the way for an “empirically grounded theory of multimodal language use and interaction” (Gregori *et al.* 2023, p. 404) – a tool as essential to the field as “the microscope to biology” (Streeck 1993, p. 275). In the digital age, where a single TikTok video or a multimodal text message combines text, emoji, filter, audio-track, and temporal editing, the ‘utterance’ is no longer a purely linguistic unit. As convincingly summarised by Grzenkowicz (2026),

¹ The introduction was jointly conceived by the three co-editors. Fabio Ciambella is the author of Section 1, Jonathan Culpeper of Section 2, and Aoife Beville of Section 3.

² “We may accompany the utterance of the words by gestures (winks, pointings, shruggings, frowns, &c.) or by ceremonial non-verbal actions. These may sometimes serve without the utterance of any words, and their importance is very obvious” (Austin 1962, p. 76).

³ The term ‘multimodality’, while already employed in fields like statistics, philosophy, and psychology, gained widespread traction in semiotics and linguistics following the publication of Kress and van Leeuwen’s foundational text, *Reading Images* (1996).

The concept of a speech act, as conceived by Searle [...] was designed to address verbal communication, focusing particularly on the lexical and grammatical dimensions of language. [...] However, speech act theory may appear defective when we, as researchers, are confronted with communication that exceeds the boundaries of a single verbal semiotic mode.

Bridging the gap between linguistic pragmatics and multimodality is a theoretical necessity because the (illocutionary and/or perlocutionary) force of a communicative act is now often non-linguistic. For example, if a content creator uses a hushing gesture while whispering, the gesture is not just supporting the word; it is performing a distinct part of the speech act – i.e., a directive in this case. Without an integrated framework, our analyses risk of remaining fragmentary, capturing only the ‘what was said’ while missing the ‘how it was meant’ and ‘how it was felt’.

First of all, the integration of these disciplines necessitates a reevaluation of the very nature of speech acts concerning how they are realised (Grzenkovicz 2026). In a multimodal framework, we propose the concept of multimodal speech acts (or communicative acts; see, among others, van Leeuwen 2004), according to which “speech acts should be renamed *communicative acts* and understood as multimodal micro events in which all the signs present combine to determine their communicative intent” (p. 121).⁴ Consider the traditional distinction between locutionary, illocutionary, and perlocutionary acts. In a multimodal environment, the locution may be, for instance, a combination of a visual image and a sound effect. Similarly, the illocution (the intent) might be realised through a specific camera angle that establishes a demand relationship with the viewer (Kress and van Leeuwen 1996, p. 133, 146).

The ‘multimodal turn’ does not so much challenge Grice’s Cooperative Principle (CP) as reveal the semiotic complexities involved in its realisation. While Grice (1975) acknowledged non-verbal examples, his primary focus remained on relatively simple, monomodal exchanges. In digital environments, however, informational cooperation is achieved through a sophisticated orchestration of modes that goes beyond the verbal. On platforms like TikTok, for instance, the Maxim of Relation (be relevant) is often satisfied via visual hashtags or background music; these cues signal the video’s genre and intent, streamlining the viewer’s cognitive processing before the creator even speaks.

Separate from this informational cooperation is the realm of social cooperation, specifically regarding politeness. Building on the work of Lakoff (1973), Leech (1983), and Brown and Levinson (1987), we can see that face-work is similarly performed through visual and spatial modes. A creator’s choice of a soft-focus filter or warm lighting, for instance, may function as a multimodal

⁴ See also Kress and Van Leeuwen (1996, p. 121) who extend Hallyday’s speech functions (1994, p. 96) to images.

positive politeness strategy, satisfying the viewer's desire to be appreciated without the need for explicit linguistic praise. Ultimately, the integration of these disciplines allows us to see that both relevance and social rapport are multimodal achievements, where non-linguistic cues set the pragmatic stage for effective communication.

Another highly influential theoretical framework for analysing multimodal pragmatics is social semiotics, particularly as advanced by van Leeuwen (2004). This perspective serves as an essential bridge between traditional linguistics and modern communication by offering a formal grammar for non-linguistic modes – treating visual, gestural, and aural elements not as mere supplements to text, but as organised systems of meaning in their own right. By strategically adapting Halliday's (1985) three metafunctions, the framework provides a robust toolkit for dissection: 1) the ideational function describes how a digital artifact represents the world and constructs experience, 2) the interpersonal (or interactive) function examines how social relations are established between creator and audience through gaze, distance, and perspective, 3) the textual (or compositional) function analyses how resources like layout and framing integrate these elements into a cohesive whole.

However, it is crucial to note that while van Leeuwen's framework effectively captures the semiotic material and structural resources used to form meaning, it does not fully capture the pragmatics of communication. In this sense, social semiotics identifies the 'potential' meanings of semiotic resources, but the actual pragmatic meaning remains dependent on how that material is interpreted within a specific, dynamic context. Through this nuanced lens, a meme or a video is viewed not just as a stream of data, but as a complex social transaction in which structured semiotic resources meet contextual inference.

Ultimately, the theoretical necessity of bridging multimodality and pragmatics can be anchored in Lawrence R. Sipe's (1998) concept of intermodal synergy. Sipe argues that in a multimodal text, the relationship between words and images – as well as gestures, audio, space, etc. – is not additive, but multiplicative; the various modes engage in a dynamic transaction that generates more complex meaning. Consequently, it is insufficient to analyse an image, a gesture, an audio clip, and a text in isolation. Instead, the focus must shift to the intermodal tension and harmony that arise during their interaction. Pragmatically, we must ask such questions as: Do the image, gesture, or intonation subvert or contradict the text to create irony or sarcasm? Do they provide elaboration by specifying a vague linguistic reference? Or do they offer enhancement, amplifying the emotional or descriptive weight of the text? Within the framework of multimodal pragmatics, these intermodal relationships function as new implicatures. Just as a listener must infer meaning from what is left unsaid in speech, a viewer must bridge the 'logic gap' between different modes. The analytical divide is finally closed when we acknowledge that the true meaning of a multimodal text does not reside within the modes themselves, but rather in the

interstices – the fertile spaces of interaction where the reader actively negotiates the synergy between the visual and the verbal.

Moving from a mono-modal to a multimodal pragmatic perspective represents more than a mere expansion of the field; it is a fundamental shift in how we define the boundaries of human interaction. By synthesising the foundational rigour of Austin, Searle, and Grice with the synergistic logics of multimodality, we arrive at a more holistic understanding of communication. This approach recognises that meaning is not a stable entity confined to the linguistic utterance, but a fluid property born from the ‘multiplicative’ interaction of diverse semiotic resources. Consequently, the study of pragmatics must look beyond the individual mode and focus on the interstices – the fertile spaces where the most profound social work of establishing rapport, performing identity, and negotiating truth is actively being done.

2. Methodological frontiers: From corpus to AI

Data is particularly important in pragmatics research. Jucker (2018, p. 3) states: “There is no research in pragmatics without data”. Even thoroughly theoretical papers usually deploy illustrative examples, i.e. data, not least in order to avoid the accusation that they are merely castles in the air. However, multimodal data presents particular challenges. Perhaps the most fundamental of these is transcription. Of course, we can do a certain amount of analysis of multimodal data in their natural environments without transcription – a video on YouTube, a tweet/post, a billboard advertisement, a pictorial novel, and so on. These types of multimodal data at least have the advantage of being relatively permanent, enabling the researcher to view and review them multiple times. Other data types are more ephemeral. One’s daily face-to-face encounters disappear into the participants’ (unreliable) memories, as do other unrecorded, real-time encounters, such as watching the performance of a play. However, all data will require transcription minimally to enable us to represent that data in publications, but also to enable systematic analyses undertaken in their own (and often lengthy) timeframe. Transcription is often associated with the writing down or typing out of spoken data. In fact, the wider, and also etymological, senses of *transcription* and *transcribe* are preferable here. The component parts, Latin *trans-* and *scribere*, are about transition and writing, not specifically from speech to writing. Indeed, the earliest source for *transcription* in the Oxford English Dictionary, John Florio’s *World of Wordes* (1558), defines it as “a transcription, a writing, a copying out” (OED, n., 1.a). Making a copy, a representation of the data, and one that is suitable for the researchers’ purposes is what is key for us.

The digital technologies that have supported the rise in multimedia communication can also support its transcription and analysis by researchers. For example, spoken words can be described following a set of conventions (e.g.

Jefferson 2004), but those conventions accommodate very little outside the oral/aural. This can be remedied by an accompanying screenshot, which could capture precise body orientations, gestures, gaze, and so forth. And of course, those spoken word transcriptions and visual screenshots need not only be a representation of real-time spoken conversation, but could, for example, also help represent various kinds of digital media (e.g. a YouTube video). This efficient and effective solution has been widely adopted. Indeed, Mondada (2018) has become something of a touchstone for multimodal transcription, and it proceeds through a combination of CA-style transcription and screenshots. This is an approach that is most suitable for qualitative analyses. The reasons for this relate to some of the limitations of screenshots. Whilst they enable rich representations, and also work well for relatively static multimodal data (e.g. a tweet/post), one is limited to what one can fit in the screenshot – they do not help us see the bigger picture (whether literally the wider visual picture or metaphorically the wider context). A screenshot would preclude, for example, any sequence of tweets/posts that extends beyond a few items. Yet pragmatics often focuses on interactional sequences that may extend to some length. Another, related limitation is that they are static: they are not optimal for dynamic behaviours and phenomena, such as those that appear in a video. Of course, one could use multiple screenshots, and this might work for in-depth qualitative analyses focused on a small stretch of data. However, things rapidly become unwieldy over longer stretches of data requiring multiple screenshots, and this is to say nothing of the fact that screenshots over dynamic data tell us nothing of what happens between the screenshots. There are, however, at least two solutions.

One solution is to supplement the kinds of conventions devised for the written transcription of oral/aural data with conventions that accommodate multiple modes. For example, in the following data fragment from a TV advertisement, the crucial thing about this spoken turn is the finger pointing by the speaker. The pointing symbol ☞ in the text below represents the beat of the wagging pointed finger.

I know this haircut is nothing to do with☞you but my☞ sister is watching outside and I want ☞her to think that ☞you did it. So ☞you sort out this terrible mess for me and I will pay you double okay you're very very nice people

The advantage of this kind of technique is that it categorises finger-pointing in the data, and an important consequence of that is that quantification becomes possible. But there are downsides. One is that inevitably there is a loss of nuance and delicacy, as is the case with all categorisation. More significantly, such a transcription could not incorporate a sufficiently wide set of transcription conventions to accommodate all that is going on in the data. Inevitably, the researcher must select certain features to focus on. The other solution, and one that has proved quite popular, is to “time-align” the actual audio or video

recording with different kinds of transcription and analysis, resulting in a tiered representation. For example, one tier could be the original video data, another tier could be the transcribed spoken words, yet another tier could be devoted to the transcription/representation of particular gestures, and so forth. The most popular tool for doing this seems to be ELAN.⁵ To illustrate, Figure 1 displays a screenshot from the Chinese TV show *Roast!*⁶ There are layers of complexity here – the host, the person roasting, the person being roasted, other participants, the studio audience, the TV audience – all contributing multimodally. On top of this, the TV audience can contribute via live-streamed commentary, seen here in the rose across the top of the screen. Transcribing and managing all this information thus represents a considerable challenge.



Figure 1
A scene from the Chinese TV show *Roast!*.

However, it can be managed within ELAN, as shown in Figure 2.

⁵ See <http://tla.mpi.nl/tools/tla-tools/elan/>.

⁶ We thank Shengnan Liu for these illustrations in Figures 1 and 2.

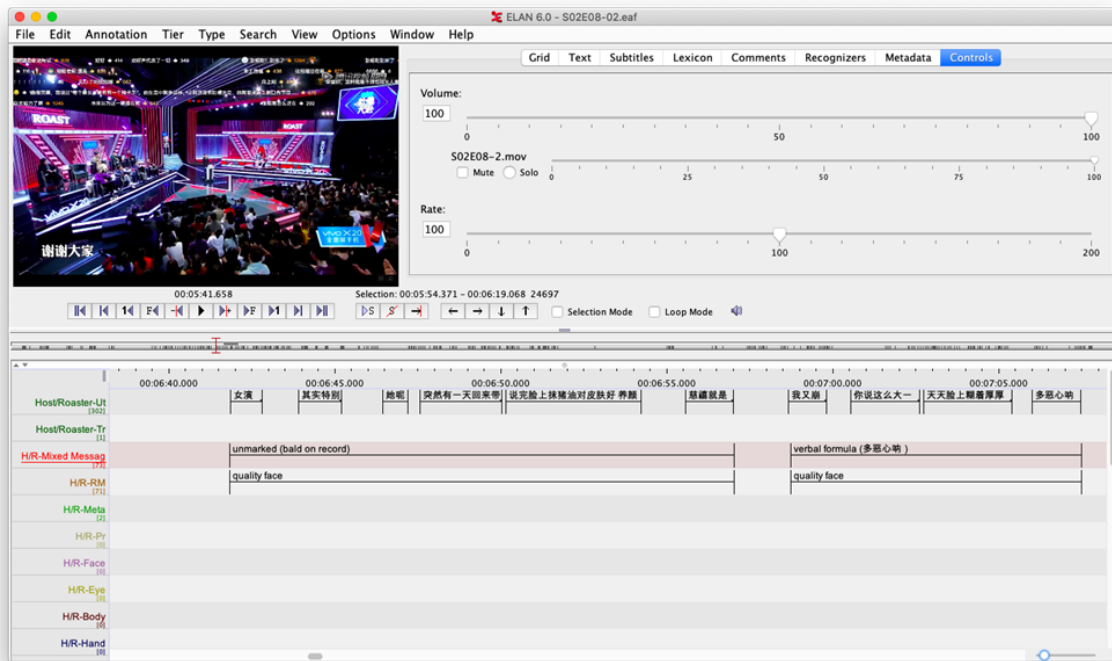


Figure 2
A scene from the Chinese TV show *Roast!* represented in ELAN.

In top left of the screenshot of the TV show in ELAN, we can see the video. The rows in the lower half of the screenshot with coloured labels to the left are crucial, as these are possible tiers of information decided on by the researcher. These could include the transcribed words, prosodic information, non-verbal behaviours such as gesture, and so forth. And importantly these tiers are time-aligned, so that when the video moves forward the information in the tiers scrolls horizontally. Such multilayered representations do not solve all problems, but they are a decisive step forward, especially where multimodal research is concerned. They allow quantification, but they also allow access to rich detail at any particular point.

One further but short step towards quantification can be taken through the development of multimodal corpora. The most common technique for creating multimodal corpora is to take the basic ideas that are reflected in ELAN and similar programs – namely, representing each mode or stream of information in a different but parallel tier or level and time-aligning them – but to do so at scale. A fairly early and well-known example of a multimodal corpus is the Nottingham Multi-Modal Corpus, a 250,000 word with aligned video, audio and textual data, using the Digital Replay System, and created between 2005 and 2008 (see Adolphs and Carter 2013).⁷ But since then the number of multimodal corpora has multiplied, often using ELAN and often quite specialist in that they are constructed to represent particular varieties of particular languages. An example

⁷ <http://thedrs.sourceforge.net/>.

is the Insight Interaction Corpus, containing spoken Flemish Dutch (see Brône, Oben 2015). This includes unscripted triadic interactions with multiple modes represented, and even eye-tracking data. What gets represented in the various tiers is of course the researchers' choice, but it can be extensive and include annotation. For example, in older corpora parts of speech would feature as codes attached to words and hidden in XML, but any kind of annotation, including codes for non-verbal aspects (e.g. gesture), can simply be represented in another tier. Whilst such corpora allow the possibility of quantitative research, it shouldn't be assumed that they are confined to that. Because they represent data in all its richness across multiple tiers, it is quite possible to analyse all the data qualitatively at one moment in time.

3. Overview of this issue

The present issue set out to bridge the gap between pragmatics and multimodality, and each contribution represents a distinct possible bridging point, both in the data analysed and in the methodological approaches employed. The issue is divided into four thematic sections, grouping together studies with similar approaches or foci.

Section 1, "Digital Media and Social Platforms", explores the multimodal and pragmatic features of digital participatory media text types, from recognisable TikTok trends to the use of emojis in digital discourse. Each paper in this section addresses a precise and timely gap in the current literature on the rapidly evolving world of digitally mediated discourse, where, naturally, the auditory, visual and gestural modes play a crucial role in the linguistic cyberscape.

Battista's contribution offers a multimodal and pragmatic analysis of *BuzzFeed's Tasty*, a veritable 'digital cookbook,' comparing approaches across platforms to the videorecipe genre in its multifaceted digital forms. In this study, multimodal pragmatic strategies (such as humour, stance, and visual/auditory stimuli) are shown to be employed to balance instruction, humour, and engagement, helping to position content creators as "expert friends" across Facebook, Instagram, and TikTok. To do so, the scholar employs the ELAN tool (as outlined in Section 2 above) to conduct a qualitative multimodal analysis, combining it with manual pragmatic annotations of illocutionary force, stance markers, and (im)politeness strategies.

Cavaliere, Corrizzato and Franceschi's paper *Emoji in R/Hikikomori. A Corpus-assisted Study of Digital Politeness, Humour, and Identity* examines the use of emojis as multimodal pragmatic devices in online-mediated discourse. Compiling a corpus of Reddit posts from the r/hikikomori subreddit, the scholars examine the multiple, context-dependent pragmatic functions of emoji use. The quantitative analysis identifies the 10 most frequently occurring emojis in the

corpus, subsequently the quantitative analysis – which seamlessly combines pragmatic and multimodal frameworks – reveals that these emojis function as multimodal pragmatic markers: expressing attitudes and emotions, signalling irony and self-deprecating humour, softening or enhancing illocutionary force, mitigating (im)politeness, and, essentially, enhancing the pragmatic weight of the digitally-mediated interactions under examination. The findings reveal the dynamic and complex role of emojis in interpersonal online discourse and in the co-construction of group and individual identities within a specific online community.

Ciambella's *Multimodal Pragmatics at Its Best: A Qualitative Analysis of ASMR Videos on TikTok* explores the strategic use of semiotic resources to promote engagement, foster intimacy, and induce a state of relaxation in viewers. In this popular format, the visual, auditory, gestural, and spatial modes are employed in order to produce specific perlocutionary effects. Outlining the specific audio and visual grammars of the format, and the characteristic features of the genre (such as whispering, close-ups, tapping sounds), Ciambella employs the category of “digital intimacy” to investigate the pragmatic import of such overlapping meaning-making processes across semiotic modes. His framework introduces the original concept of “multimodal pragmatic (mis)alignment”, whereby the modes produce coherent (aligned) or contrasting (misaligned) illocutionary or perlocutionary outcomes: in the first case, heightening effective resonance; in the second, producing irony. The study also examines the (im)politeness strategies at play in the genre, focusing specifically on “pseudo-rogative speech acts” as pragmatic strategies intended to enhance the simulation of interpersonal interaction and promote a sense of intimacy.

Similarly, Willoughby's study examines the pragmatic and multimodal features of a popular TikTok format, specifically the POV (point-of-view) structure. Drawing from humour studies in pragmatics and linguistics, she investigates “cross-modal incongruity” as a unique type of modal interplay, whereby one semiotic mode's cues disrupt or shift the interpretive frame established by another mode's cues. Through rigorous data segmentation and analysis, the study reveals the centrality of cross-modal misalignment in the dataset, showing how content creators engage multiple semiotic resources to convey incongruity for humorous effects. The findings reveal, therefore, a genre-specific strategy that relies on pragmatic and multimodal approaches (both delivery and uptake) to be successful.

The second section of the issue focuses on the use of Artificial Intelligence in educational settings. Drawing insights from language teaching theory and practice, multimodal critical discourse analysis, and pragmatic theories of conversational interaction, this section provides two contrasting, yet complementary, perspectives on student-machine interfacing. While with appropriate instructor involvement and supervision, AI tools can be beneficial in

EFL contexts, their current lack of pragmatic and multimodal competence poses significant risks in managing nuanced and volatile discursive practices.

Giacosa's contribution, *AI-powered Multimodal Feedback in EFL Writing: Pragmatic Functions and Classroom Implications*, examines the practical pedagogical applications of AI feedback for EFL students, with a specific focus on the multimodal and pragmatic features of the AI-driven automated written corrective feedback (AWCF) tool Grammarly. By analysing the use of visual and textual semiotic resources, the study examines how AI-mediated feedback may reshape the pragmatic dynamics of the EFL classroom. It first investigates the illocutionary force of the tools' corrections and suggestions and subsequently explores their perlocutionary effects on student motivation and revision strategies through learner questionnaires. Finally, by integrating these findings with EFL teachers' self-evaluations of multimodal feedback strategies, this study explores the potential benefits of using such tools under teacher supervision to foster student-centred practices and enhance instructors' classroom interactional competence.

Pizzo's essay, *Representations of Hate Discourse: Interactions Between Human Linguistic Pragmatics and AI Multimodal Pragmatics*, similarly uses data from students' engagement with AI; however, in this study, the focus shifts to the interactional dynamics of the student-machine conversations, as they ask generative AI models, such as ChatGPT or Copilot, to produce images related to hate speech. In the first phase of the research, the student-generated prompts and subsequent AI responses are analysed according to Grice's Cooperative Principle, revealing complex implicatures regarding religious identity, ethnicity, gender, and body image. Pizzo's findings reveal that the AI models used do not grasp the sociopragmatic nuances of human identity and intersectionality, often resorting to stereotypes. In the second phase of the research undertaken, the analysis focuses on the final image output from the exchange. The study brings to light that the AI models analysed rely too heavily on the linguistic mode; visual modality does not emerge as semiotically autonomous. Furthermore, the findings suggest that students may defer to AI's perceived "epistemic authority", even on nuanced topics such as the discourse of hate. Such findings raise ethical concerns about AI data biases and its inability to engage appropriately with hate discourse, a complex and multifaceted phenomenon of human interaction.

The third section of the issue focuses on the pragmatic and multimodal analysis of performative and aesthetic texts, from radio drama and stage theatre to film and the visual arts. This section demonstrates the utility of stylistic methodologies and approaches for aligning pragmatic and multimodal frameworks of analysis and reveals the complexity of applying such models to fictional and aesthetic texts, which function both on an inter-character (intradiegetic) and a character-audience (extradiegetic) level.

Beville's – admittedly provocatively titled – *The Hypostatic Union of Multimodality and Pragmatics: Exploring the Auditory Mode in Dorothy L.*

Sayers' Radio Plays 'The Man Born To Be King', borrows the term 'hypostatic union' from Christian theology in order to explore the points of contact between the auditory and the linguistic semiotic modes. The qualitative multimodal pragmastylistic analysis examines D.L. Sayers' cycle of audio plays based on the biblical Gospel narratives and the representation of characters through overlapping and interconnected auditory and linguistic modes. The qualitative, stylistic approach to the auditory and pragmatic features of the selected extracts reveals the process of constructing the plays' characters, specifically the multidimensional representation of the central figure, Jesus, as both human and divine. Furthermore, the paper encourages a deeper understanding of the bridging points between pragmatics and multimodality: ostensibly auditory features, such as accent, tone, prosody, and sound effects, are shown to alter pragmatic dynamics at both intradiegetic and extradiegetic levels. Similarly, sociopragmatic features, such as (im)politeness and facework, interactional storytelling, and features of sociolinguistic variation, are shown to be *voiced* through the auditory mode and utterly dependent on the voice actors' embodied performances.

Ghezzi analyses Shakespeare's Ophelia across various semiotic representations which have followed from her fleeting appearance in *Hamlet*. Her intersemiotic pragmastylistic analysis aims to demonstrate how pragmatic and aesthetic transformations through transmediation can affect characterisation and highlight features of source and target texts. Engaging with Ophelia's linguistic and visual representations in key excerpts of the Shakespearean play, a film adaptation (McCarthy's *Ophelia*, 2018), and two famous pictorial representations (portraits by Waterhouse and Millais), the study reveals how such intersemiotic translations can reinforce or challenge the schemata that underlie interpretations of the character.

Vetri's contribution, *Multimodal Pragmatics and the Stylistics of Drama: Meaning-Making in Performance in David Hare's Skylight*, illustrates the harmonious and successful integration of pragmatic and multimodal analysis within a stylistic framework. By analysing a recorded theatrical performance as embodied discourse, the scholar demonstrates how prosody, kinesics, and proxemics serve as semiotic resources that actors strategically employ alongside verbal utterances. The multilayered analysis reveals the complexity of the relationship at the core of the play, which enacts interpersonal conflict, emotionally-charged tension, and ideological battles. *Skylight*, specifically, and theatre more generally, cannot be fully understood through a single mode. The semiotics of theatre illustrates how meaning is conveyed through gesture, space, and paralinguistic resources, while pragmatics uncovers the underlying communicative intent that fuels interpersonal interactions. Together, these perspectives demonstrate that the characters' relationship fails not just because of ideological divergence but also because of their mutual communicative inflexibility: they are trapped in a conflict in which both sides pursue victory over mutual understanding.

The fourth and final section of this issue deals with pragmatics and multimodality in intercultural, interpersonal, and corporate discourse, examining thank-you cards, English-as-a-second-language interactions, and corporate social responsibility statements. Each paper seeks to bridge interpersonal, intercultural and intersemiotic gaps by applying multimodal pragmatic frameworks of analysis. Moving from the study of intimate exchanges (thank you cards), to intercultural dynamics (Egyptian English-mediated interactions) and finally to institutional discourse (ESG and CSR statements).

Bertoldi and Gesuato examine the intricate interplay between the verbal and visual modes in the macro-genre of commercially printed thank-you cards. The scholars analyse a corpus of over 90 thank-you cards with an integrated model of speech act analysis, social semiotics, and intersemiotic relations. This framework reveals the representational structures and compositional arrangement of the cards, as well as the complex interplay between the mode-specific affordances of the linguistic and visual modes, which overlap through their co-deployment in thanking acts mediated through the commercial thank-you card. Their findings reveal that strategies such as typographical prominence, concise, explicit expressions of gratitude, and visually anchored humour are used to engage audiences (firstly, the consumer-thanker, and secondly, the thankee). Furthermore, politeness facework is shown to operate through intersemiotic processes of meaning-making. While verbal elements prioritise clarity and immediacy, the visual aspects serve relational and affective goals through humour and aesthetic enhancement.

La Causa's contribution analyses the intercultural aspects of multimodal communication, examining a case study of an Egyptian speaker's English-mediated interactions. Multimodal cues, such as prosody, proxemics, oculesics, haptics and kinesics, are culturally dependent; therefore, cross-cultural communication presents complex challenges for speakers and their audiences. Through her study of Egyptian actor and satirist Bassem Youssef's interactions in English-mediated contexts, the scholar reveals some of the possible multimodal pragmatic misfires, failures, and misinterpretations that can occur in such intercultural exchanges. The same multimodal pragmatic features that may be perceived as warmth or solidarity by one audience may be seen as aggression, intrusion, or exaggeration by another audience with contrasting sociopragmatic expectations regarding gesture, proximity, touch, and prosody.

Zaupa's paper focuses on the verbal and visual representations of diversity, equity, and inclusion (DEI) in a corpus of corporate discourse, including CSR (corporate social responsibility) and ESG (environmental, social, and governance) disclosures. His findings indicate a focus on corporate efforts to promote equality for LGBTQ+ employees and enhance the social well-being of local LGBTQ+ communities where companies operate. Although this information is ostensibly disclosed transparently (complete, accurate, and relevant), linguistic vagueness may diminish the overall clarity and transparency.

The analysis of the visual components reveals an intent to promote visibility of certain sexual and gender identities, but the balanced portrayal of specific identities alongside generic individuals challenges diversity narratives. While the factual tone and visual focus on diverse identities suggest genuine corporate involvement, the vague language indicates a communication strategy that is partly constructed and possibly opportunistic.

Ultimately, across its four sections, this issue of *Lingue e Linguaggi*, successfully addresses the theoretical and empirical gap identified at its inception, moving from, and beyond, traditional pragmatics towards a broader understanding of discursive, situated, multimodal, communicative events. By bringing together scholars across digital, educational, aesthetic, intercultural, and corporate domains, this collection of essays achieves its primary aim: to demonstrate that pragmatic meaning is not merely supplemented by non-verbal modes; rather, it is necessarily constituted through various interconnected and overlapping semiotic resources. Stylistics, CDA, corpus linguistics, and humour studies, variously, serve as vital bridge disciplines, allowing scholars to explore the complex and nuanced landscapes of interactional and intersemiotic communication, even beyond the solely ‘human’ connotations of such terms. In fact, this issue further pushes the boundary of the once revolutionary statement that “multimodality [is] the normal state of human communication” (Kress, 2010 p. 1); on the one hand, we hope to have shown that *pragmatic* multimodality is the normal state of human communication, on the other hand, we can not ignore *non-human* communication in our rapidly evolving AI- and digitally-enhanced world.

Bionotes: Aoife Beville obtained her PhD from the University of Naples *L’Orientale* in 2022. She is a Researcher in English Linguistics at the University of Foggia and the author of *Telling Tales in Shakespeare’s Drama: A Pragma-Stylistic Approach to Lying* (Paolo Loffredo Editore, 2022). While she continues to explore the pragmatics of mendacity in a variety of texts (films, false confessions, garment labels, etc.), her recent publications also focus on the stylistics of persuasion in religious discourses and the functions of irony in poetry. She is the administrative coordinator of the Argo Research Centre (*Centro di Ricerca Interuniversitario di Argomentazione, Pragmatica e Stilistica*), where she also collaborates in directing research projects on ecostylistics, (im)politeness in cinema, and the stylistics of *pseudos*.

Fabio Ciambella is a Researcher of English Language, Linguistics and Translation at Sapienza University of Rome. His research interests include the lexicography of dance in Early Modern England, historical pragmatics, Second Language Acquisition, and culinary linguistics. His most recent publications are *Dance Lexicon in Shakespeare and His Contemporaries: A Corpus-based Approach* (Routledge, 2021), and *Teaching English as a Second Language with Shakespeare* (Cambridge University Press, 2024). He is a member of two research projects of relevant national interest (PRIN): PoWoR (Politics of Worship Pre-and Post-Reformation) and SENS (Shakespeare’s Narrative Sources: Italian Novellas and Their European Dissemination). Jonathan Culpeper is Professor of English Language and Linguistics at Lancaster University, UK. His research combines pragmatics – especially interpersonal pragmatics and

(im)politeness theory – with historical linguistics, focusing on the history of English and Early Modern English, and with stylistics, particularly the language of Shakespeare and drama. He often employs corpus-based and computational methods to explore how language shapes social interaction across time. He has authored and (co-)edited numerous influential works, including *Impoliteness: Using Language to Cause Offence* (Cambridge University Press, 2011), *Pragmatics and the English Language* (with Michael Haugh, Palgrave Macmillan, 2014), and *Second Language Pragmatics: From Theory to Research* (with Alison Mackey and Naoko Taguchi, Routledge, 2018). His work has made major contributions to linking linguistic theory with literary and historical language study, and he is widely recognised as one of the foremost scholars in (im)politeness research and historical pragmatics.

Authors' addresses: aoife.beville@unifg.it; fabio.ciambella@uniroma1.it; j.culpeper@lancaster.ac.uk

Acknowledgements: We would like to thank all the contributors to this issue for their patience, attention to detail, and accuracy. Furthermore, we thank the anonymous peer reviewers of each contribution for their helpful insights.

References

- Adolphs S. and Carter R. 2013, *Spoken Corpus Linguistics: From Monomodal to Multimodal*, Routledge, London.
- Austin J.L. 1962, *How to Do Things with Words*, Oxford University Press, Oxford.
- Brône G. and Oben B. 2015, *InSight Interaction: A Multimodal and Multifocal Dialogue Corpus*, in “Language Resources and Evaluation” 49, pp. 195-214.
- Gregori A. et al. 2023, *A Roadmap for Technological Innovation in Multimodal Communication Research*, in Duffy V.G. (ed.), *Digital Human Modeling and Applications in Health, Safety, Ergonomics and Risk Management: 14th International Conference, DHM 2023, Held as Part of the 25th HCI International Conference, HCII 2023, Copenhagen, Denmark, July 23-28, 2023, Proceedings, Part II*, Springer, Cham, pp. 402-438.
- Grice H.P. 1975, *Logic and Conversation*, in Davidson D. (ed.), *The Logic of Grammar*, Dickenson Publishing Company, Encino (CA), pp. 64-75.
- Grzenkiewicz M. 2026, *Non-verbal Artifacts and Propositionality: Adjusting Speech Act Theory to Accommodate Multimodal Argumentation*, in “Argumentation: An International Journal on Reasoning” 40 [1], n.p.
- Halliday M.A.K. 1985, *Spoken and Written Language*, Oxford University Press, Oxford.
- Halliday M.A.K. 1994, *An Introduction to Functional Grammar*, Edward Arnold, London.
- Jefferson G. 2004, *Glossary of Transcript Symbols with an Introduction*, in Lerner G.H. (ed.), *Conversation Analysis: Studies from the First Generation*, John Benjamins, Amsterdam, pp. 13-31.
- Jucker A.H. 2018, *Data in Pragmatic Research*, in Jucker A.H., Schneider K.P. and Bublitz W. (ed.), *Methods in Pragmatics*, vol. 10, Walter de Gruyter, Berlin, pp. 3-36.
- Kress G.R. and van Leeuwen T. 1996, *Reading Images: The Grammar of Visual Design*, Routledge, London/New York.
- Kress G.R. 2010, *Multimodality. A Social Semiotic Approach to Contemporary Communication*, Routledge, London/New York.
- Lakoff R.T. 1973, *The Logic of Politeness; or, Minding your P's and Q's*, in Corum C., Cedric Smith-Stark T. and Weiser A. (ed.), *Papers from the Ninth Regional Meeting of the Chicago Linguistics Society*, University of Chicago Press, Chicago (IL), pp. 292-305.
- Leech G.N. 1983, *Principles of Pragmatics*, Longman, London.
- Mondada, L. 2018, *Multiple Temporalities of Language and Body in Interaction: Challenges for Transcribing Multimodality*, in “Research on Language and Social Interaction” 51 [1], pp. 85-10.
- Searle J. 1969, *Speech Acts: An Essay in the Philosophy of Language*, Cambridge University Press, Cambridge.
- Sipe L.R. 1998, *How Picture Books Work: A Semiotically Framed Theory of Text-Picture Relationships*, in “Children’s Literature in Education” 29 [2], pp. 97-108.
- van Leeuwen T. 2004, *Introducing Social Semiotics*, Routledge, London/New York.