

The present issue of *Lingue e Linguaggi* sets 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. Across its four sections, it 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.

Chapters by:

Aoife Beville, Fabio Ciambella, Jonathan Culpeper

Alessia Battista

Silvia Cavaliere, Valeria Franceschi, Sara Corizzato

Fabio Ciambella

Audrey Claire Willoughby

Antonella Giacosa

Raffaele Pizzo

Aoife Beville

Chiara Ghezzi

Valentina Vetri

Elisa Bertoldi, Sara Gesuato

Lucia La Causa

Federico Zaupa

Lingue e Linguaggi

vol. 74 - Special Issue 2026



Bridging the Gap between Pragmatics and Multimodality

Edited by

Aoife Beville

Fabio Ciambella

Jonathan Culpeper

Lingue e Linguaggi

vol. 74 - Special Issue
2026



Università del Salento

Lingue & Linguaggi

74/2026

Special Issue

Bridging the Gap between Pragmatics and Multimodality

Edited by
Aofie Beville
Fabio Ciambella
Jonathan Culpeper



UNIVERSITÀ
DEL SALENTO

2026

LINGUE E LINGUAGGI

Pubblicazione del Dipartimento di Studi Umanistici dell'Università del Salento. Tutti i contributi pubblicati in *Lingue e Linguaggi* sono stati sottoposti a *double-blind peer-review*.

Numero 74/2026

COMITATO DI CONSULENZA SCIENTIFICA DELLA RIVISTA

Orietta Abbati, Università di Torino
Jörn C. Albrecht, Rupprecht-Karls-Universität Heidelberg
Pedro Álvarez de Miranda, Real Academia Española
Carmen Argondizzo, Università della Calabria
Sara Augusto, Universidade de Coimbra
Gabriele Azzaro, Università di Bologna
Marcos Bagno, Universidade de Brasília
Carla Barbosa Moreira, CEFET- MG - Brasile, Brazil
Jean-Philippe Barnabé, Université de Picardie (Amiens, Francia), France
Simona Bertacco, University of Louisville, USA
Silvia Betti, Alma Mater-Università di Bologna e ANLE (Academia Norteamericana de la Lengua Esp.), Italy
Francesco Bianco, Univerzita Palackého v Olomouci
Alessandro Bitonti, Masaryk University, Czech Republic
Jean-Claude Blachère, Université Paul Valéry, Montpellier III
Maria Bortoluzzi, Università degli Studi di Udine
Chiara Bolognese, Universidad Autónoma de Barcelona
Lorella Bosco, Università degli Studi di Bari
Maria Luiza Braga, Universidade Federal do Rio de Janeiro
Cristiano Broccias, Università degli Studi di Genova
Silvia Bruti, Università degli Studi di Pisa
Sandra Campagna, Università degli Studi di Torino
Catherine Camugli Gallardo, Université Paris Ouest - Nanterre
Xelo Candel Vila, Universitat de València
Martha Canfield, Università degli Studi di Firenze
Gloria Cappelli, Università di Pisa, Italy
Manuel Carrera Díaz, Universidad de Sevilla
Alessandro Cassol, Università degli Studi di Milano
Gabriella Catalano, Università di Roma "Tor Vergata"
Paola Catenaccio, Università degli Studi di Milano
Marco Cipolloni, Università di Modena e Reggio Emilia
Carmen Concilio, Università degli Studi di Torino
Alessandro Costantini, Università degli Studi di Venezia
Pier Luigi Crovetto, Università degli Studi di Genova
Maria Debora De Fazio, Università degli Studi della Basilicata
Giorgio de Marchis, Università Roma Tre
Maria del Valle Ojeda Calvo, Università degli Studi di Venezia "Ca' Foscari"
Arthur DeFrance, Università di Waseda, Tokyo
Jean-Christophe Delmeule, Université Charles De Gaulle, Lille 3
Gabriella Di Martino, Università degli Studi di Napoli
Marina Dossena, Università degli Studi di Bergamo, Italy
Jean-François Durand, Université Paul Valéry, Montpellier III
Claus Ehrhardt, Università degli Studi di Urbino, Italy
Roberta Facchinetti, Università degli Studi di Verona
Federica Ferrari, Università degli Studi di Bologna
Teresa Ferrer Valls, Universitat de València
Luisanna Fodde, Università degli Studi di Cagliari
Giuliana Elena Garzone, Libera Università IULM di Lingue e Comunicazione
Sara Gesuato, Università degli Studi di Padova
Barbara Gori, Università degli Studi di Padova
Enrico Grazzi, Università degli Studi Roma 3
Dorothee Heller, Università degli Studi di Bergamo
Laetitia Jensen Eble, Universidade de Brasília
Mersini Karagevrekci, University of Macedonia
Jean René Klein, Université catholique de Louvain

Emil Lafe, Centro di Studi Albanologici, Tirana
Elena Landone, Università di Sassari
Anna Maria Laserra, Università degli Studi di Salerno
Lucilla Lopriore, Università degli Studi Roma 3, Italy
Monica Lupetti, Università di Pisa
Stefania Maci, Università degli Studi di Bergamo
Aldo Antonio Magagnino, Professional literary translator, Italy
Francisco Martín, Università degli Studi di Torino
Antonio Manieri, Università di Napoli "L'Orientale"
Federico Masini, "Sapienza" Università di Roma
Daniela Mauri, Università degli Studi di Milano
Selena Millares, Universidad Autónoma de Madrid
Sandro M. Moraldo, Università di Bologna
Rafael Morales Barba, Universidad Autónoma de Madrid, Spain
Marina Morbiducci, Università di Roma Sapienza
Mara Morelli, Università degli Studi di Genova
Francesco Morleo, Italy
Martina Nied, Università di Roma Tre
Liana Nissim, Università degli Studi di Milano
Vincenzo Orioles, Università degli Studi di Udine
Elisa Perego, Università degli Studi di Trieste, Italy
Alain Peyraube, Ecole des Hautes Etudes en Sciences Sociales (EHESS, Paris) and Emeritus director of research at the Centre National de la Recherche Scientifique (CNRS, Paris, France)
Barbara Sasse, Università degli Studi di Bari "Aldo Moro"
Francesco Saverio Perillo, Università degli Studi di Bari
Ronny F. Schulz, Christian-Albrechts-Universität su Kiel
Elena Pessini, Università degli Studi di Parma
Diane Ponterotto, University of Rome Tor Vergata
Franca poppi, Università di Modena e Reggio Emilia
Chiara Preite, Univ. di Modena e Reggio Emilia
Virginia Pulcini, Università di Torino
Lupe Romero Ramos, Universidad Autónoma de Barcelona
Alessandra Riccardi, Università di Trieste
Silvia Riva, Università degli Studi di Milano
Federica Rocco, Università degli Studi di Udine
José-Carlos Rovira Soler, Universidad de Alicante
Mette Rudvin, Università di Bologna, Italy
Vincenzo Russo, Università di Milano
Rita Salvi, Università di Roma "La Sapienza"
Antonio Sánchez Jiménez, Université de Neuchâtel
Julián Sauquillo Gómez, Universidad Autónoma de Madrid
Michael Schreiber, Johannes-Gutenberg-Universität Mainz
Marcello Soffritti, Università degli Studi di Bologna
Elena Spandri, Università degli Studi di Siena
Valeria Tocco, Università di Pisa
Ilda Tomas, Università di Granada, Spain
Georgina Torello, Universidad de la República (Montevideo)
Nicoletta Vasta, Università di Udine
Germán Vega García-Luengos, Universidad de Valladolid
Ivan Verc, Università degli Studi di Trieste
Graciela Villanueva, Université de Paris Est Créteil Val de Marne
Itala Vivan, Università degli Studi di Milano
Bryan Weston Wyly, Università della Val D'Aosta
Raúl Zamorano Farias, Universidad Nacional Autónoma de México

DIRETTORE RESPONSABILE: Maria Grazia Guido, Università del Salento

DIRETTORE DEL DIPARTIMENTO DI STUDI UMANISTICI E DIRETTORE SCIENTIFICO: Maria Grazia Guido, Università del Salento

JOURNAL MANAGER: Pietro Luigi Iaia, Università del Salento

COMITATO EDITORIALE: Marcello Aprile, Yasemin Bayyurt, Miguel Bernal-Merino, Francesca Bianchi, Gualtiero Boaglio, Victoria Bogushevskaya, Thomas Christiansen, Alessia Cogo, Rosita D'Amora, Giulia D'Andrea, Antonella De Laurentiis, Maria Luisa De Rinaldis, Gian Luigi De Rosa, Martin Dewey, Jorge Díaz Cintas, Giulia Andreina Disanto, Maria Renata Dolce, Monica Genesin, Maria Teresa Giampaolo, Barbara Gili Fivela, Mirko Grimaldi, Maria Grazia Guido, Gerhard Hempel, Pietro Luigi Iaia, Marcella Leopizzi, Elena Manca, Simona Mercantini, Maria Chiara Migliore, Fabio Moliterni, Antonio Montinaro, Rocco Luigi Nichil, Gloria Politi, Luciano Ponzio, Mariarosaria Provenzano, Irene Romera Pintor, Amalia Rodríguez Somolinos, Alessandra Rollo, Andrea Scardicchio, Virginia Sciutto, Diego Simini, Carmelo Spadola, Anne Theissen, Elena I. Trofimova, Natasha Tsantila.

DIREZIONE E REDAZIONE

Dipartimento di Studi Umanistici
73100 LECCE, via Taranto, 35
tel. +39-(0)832-294401, fax +39-(0)832-249427
Copertina di Luciano Ponzio: *Ecriture* (particolare), 2007.

© 2026 University of Salento - Coordinamento SIBA

<http://siba.unisalento.it>

ISSN 2239-0367

eISSN 2239-0359 (electronic version)

<http://siba-ese.unisalento.it>



Indice

- 5 AOIFE BEVILLE, FABIO CIAMBELLA, JONATHAN CULPEPER, *Introduction: Bridging the Gap between Pragmatics and Multimodality*

SECTION 1: DIGITAL MEDIA AND SOCIAL PLATFORMS

- 21 ALESSIA BATTISTA, *Positioning the 'Expert-Friend' in the Digital Kitchen. A Multimodal-Pragmatic Analysis*
- 43 SILVIA CAVALIERI, VALERIA FRANCESCHI, SARA CORRIZZATO, *Emoji in R/Hikikomori: A Corpus-Assisted Study of Digital Politeness, Humour, and Identity*
- 69 FABIO CIAMBELLA, *Multimodal Pragmatics at Its Best: A Qualitative Analysis of ASMR Videos on TikTok*
- 87 AUDREY CLAIRE WILLOUGHBY, *Humour and Modal Interplay in TikTok's POV Genre*

SECTION 2: ARTIFICIAL INTELLIGENCE AND EDUCATIONAL TECHNOLOGY

- 107 ANTONELLA GIACOSA, *AI-powered Multimodal Feedback in EFL Writing: Pragmatic Functions and Classroom Implications*
- 135 RAFFAELE PIZZO, *Representations of Hate Discourses: Interactions between Human Linguistic Pragmatics and AI Multimodal Pragmatics*

SECTION 3: MULTIMODAL PRAGMATICS IN FICTIONAL AND AESTHETIC CONTEXTS

- 173 AOIFE BEVILLE, *The Hypostatic Union of Multimodality and Pragmatics: Exploring the Auditory Mode in Dorothy L. Sayers' Radio Plays The Man Born to Be King*
- 195 CHIARA GHEZZI, *Ophelia's Character between Imagery and Interaction: An Intersemiotic Pragma-stylistic Analysis*
- 223 VALENTINA VETRI, *Multimodal Pragmatics and the Stylistics of Drama: Meaning-Making in Performance in David Hare's Skylight*

SECTION 4: SOCIAL RESPONSIBILITY AND INTERCULTURAL INTERACTION

- 245 ELISA BERTOLDI, SARA GESUATO, *The Multimodal Expression of Gratitude: The Interplay of Words and Pictures in Thank-you Cards*
- 275 LUCIA LA CAUSA, *Multimodal Pragmatics in Intercultural Communication: Egyptian Multimodal Cues in English-mediated Interactions*
- 301 FEDERICO ZAUPA, *Sincere or Opportunistic Corporate Inclusion? A Pragmatic and Multimodal Corpus-informed Approach to LGBTQ+ (In/Ex)clusion in CSR and ESG Disclosures*

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 Grzenkiewicz (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 (Grzenkiewicz 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 *Worlde 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.

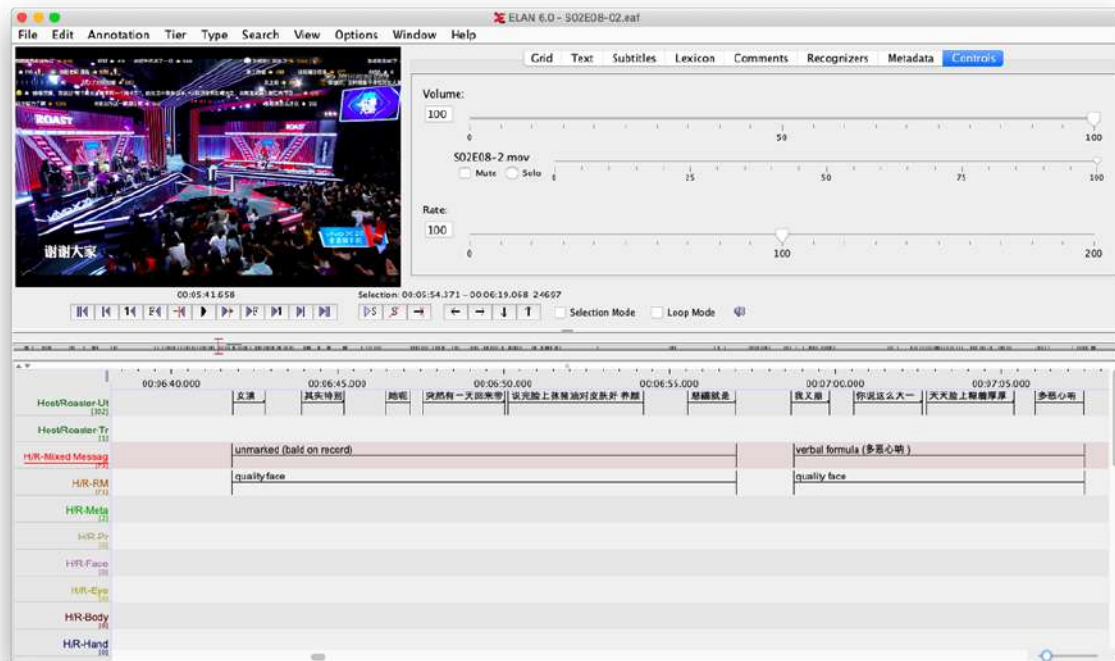


Figure 2

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.

Cavalieri, 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.

POSITIONING THE ‘EXPERT-FRIEND’ IN THE DIGITAL KITCHEN A Multimodal-Pragmatic Analysis

ALESSIA BATTISTA
TOR VERGATA UNIVERSITY OF ROME

Abstract – This contribution presents a multimodal-pragmatic analysis of *BuzzFeed’s Tasty*, a US-American digital food media brand, launched in 2015 and rapidly grown into a brand with a strong social media presence. *Tasty* can be considered as a ‘digital cookbook’ revitalising the traditional ‘recipe book’ genre through multimodal formats and multisensory content, which is a successful strategy to attract Gen-Z and Millennials (Battista 2025b). Considering a sample of videos posted in 2023 on Facebook, Instagram, and TikTok, this study examines how *Tasty* orchestrates verbal and non-verbal elements to construct meaning, focusing on interpersonal pragmatics to explore the ‘expert-friend’ persona (Battista 2025a). The methodology adopts a cross-platform perspective (Brunner and Diemer 2022) and integrates multimodality (Kress and van Leeuwen 2020), (im)politeness strategies (Culpeper *et al.* 2017), stancetaking (Kiesling 2022), and speech act theory (Austin 1962). Findings reveal that *Tasty* creators construct an ‘expert-friend’ persona blending expertise and relatability through humour, softened directives, self-disclosure and self-mockery. The cross-platform approach shows that Facebook videos foreground narrative empathy and playfulness, Instagram videos rely on concise and visually driven instructions, while TikTok prioritises personal storytelling and cultural/traditional references. Across all platforms, multimodal cues reinforce clarity and engagement, pragmatically intensifying instruction and affiliation, thus shaping user engagement and trust in digital food discourse.

Keywords: recipe video; social media; pragmatics; multimodality.

1. Introduction and background

Social media platforms have transformed how people create, access, and consume content, with short-form videos becoming popular over the last decade (Zhang *et al.* 2022). This is particularly evident in the domain of food and cooking, where social media platforms have contributed to the transformation of recipe videos into multisensory interactive digital cookbooks, whose features include catchy music, on-screen text, and engaging the audience (Battista 2025a). Short and engaging videos permeate social media; however, systematic frameworks for their unique communicative and aesthetic features remain limited (Wang 2021), and more research on their

formal and aesthetic attributes is needed (Grzenkiewicz and Wildfeuer 2025; Schluskel and Frosh 2023). While studies on multimodal meaning-making in digital environments exist (Adami and Jewitt 2016; Bouvier and Machin 2018; Zappavigna 2015), research on the pragmatic value of multimodal resources in social media videos remains scarce. In particular, little is known about the use of multimodal strategies in social media videos for persuasive, humorous, or instructional goals (for a recent, comprehensive overview of the issue, see Jucker 2024).

This paper explores meaning-making processes on social media by considering *BuzzFeed's Tasty* as a case study and relying on well-established linguistics and pragmatics theories. More specifically, this paper investigates the use of multimodal pragmatic strategies to balance instruction, humour, and engagement across three social media platforms, namely Facebook, Instagram, and TikTok.

Theories in discourse and genre studies have provided scholars with frameworks originally conceived for the analysis of communication in traditional contexts, such as print news, academic writing, and face-to-face conversations (Bhatia 2008; Kress and van Leeuwen 2020). Nevertheless, the constant and rapid evolution of genres in the digital age poses some challenges. Social media content is inherently multimodal, blending speech, text, visuals, and sounds, while also seeking to engage users and to interact with them in more or less direct ways (for different perspectives on the topics of social media content and interaction, see Ashley and Tuten 2015; Cunningham and Craig 2021; Kang *et al.* 2020). This explains how once familiar genres have changed under the influence of digital formats and social media, serving their purposes but in more complex ways, mediating form, content, action, and medium (Miller 2015). As a consequence, online genres such as recipe videos cannot be comprehensively analysed by focusing on text only. Multiple semiotic modes should be considered at once, as their interaction contributes to the genre's social action (Austin 1962; Kress and van Leeuwen 2001, 2020; Miller 2015). This perspective aligns with recent developments in multimodal pragmatics, which view speech, gestures, images, and sounds as interconnected semiotic resources and highlight that illocutionary force is not limited to verbal utterances but involves multiple semiotic modes (Payrató 2017; O'Halloran *et al.* 2014). Nonetheless, discourse theories need to be recalibrated to fully adapt to the distinctive features and participatory nature of social media.

Recipe videos can be an interesting case study in this sense. Recipes are an ancient genre, dating back to the II Millennium BC (Arendholz *et al.* 2013; Notaker 2024), and can be described as stable and classic, since its main goal (instructing) and structure (Arendholz *et al.* 2013; Garzone 2017; Görlach 1992) have remained unaltered. However, online forms of communication have impacted them, leading to food blogs (Cesiri 2020), branded webpages

and social media accounts, such as *BuzzFeed's Tasty*. Launched in 2015 as a Facebook page, *Tasty* pioneered the so-called 'top-down' videos: short, overhead-style recipes, with upbeat music and text overlays for instructions. This format allowed the brand to retain the instructional goal of a recipe while adapting it to the quick pace of social media (Battista 2025b). Then, with the advent of other platforms, such as Instagram and TikTok, *Tasty's* videos have become more elaborate, with 'food-porn' and ASMR/satisfying aesthetics (Del Campo and Kehle 2016; McDonnell 2016), occasionally presenting food-themed challenges alongside recipes, resulting in a hybrid *infotainment* discourse (Vidgen 2016). This hybridisation reflects broader trends in online discourses, where traditional genres are constantly reshaped through participatory affordances and because of the conventions that emerge around them (Abidin 2018; Baldauf *et al.* 2017; Pérez-Llantada 2024). Furthermore, social media recipe videos entertain the audience through storytelling, humour, and informal conversational tones, while still being instructive; one of the distinctive features of *Tasty's* videos is that of the 'expert friend' persona, which is pragmatically constructed through stancetaking and politeness strategies and blends culinary expertise with informality, friendliness, and accessibility (Battista 2025a). This persona is not merely linguistic; it is pragmatically and multimodally constructed, which resonates with other studies suggesting that authority blends expertise and relatability and is multimodally negotiated (Abidin 2018). Therefore, this article proposes that one of the most successful communicative strategies used by *Tasty* is precisely the use of multimodal resources to perform pragmatic functions contributing to both the instructional and engaging goals of its videos. The 'expert friend' persona aligns with current views of interpersonal pragmatics and with stancetaking and relational engagement seen as central components of digital identity construction (Kiesling 2022). Through multimodal stance cues, such as evaluative language, direct address, and humour, the creator maintains both epistemic authority and affiliative proximity, which can benefit engagement. By considering a selection of videos posted in 2023 by the brand, this study shows how pragmatic theories – namely speech acts (Austin 1962), terminology (Cesiri 2016), stancetaking (Kiesling 2022; Kiesling *et al.* 2018), (im)politeness (Andersson 2024; Culpeper *et al.* 2017; Turkle 2015) – can be used to analyse multimodal digital content while bridging the gap between pragmatics and multimodality. These frameworks enable the analysis of pragmatic functions in multimodal digital contexts and contribute to the conceptualisation of how digital genres adapt to new participatory and sensory environments (Sanfilippo 2023; Schluskel and Frosh 2023). Furthermore, the communicative strategies used by creators are shaped by each platform's affordances, which also influence audience engagement (Bouvier and Machin 2018). Overall, this study contributes to the still underexplored field of research on multimodal pragmatics in digital contexts, in which meaning is dynamically

constructed through the interaction of verbal and non-verbal, linguistic and paralinguistic semiotic resources. By combining established pragmatic frameworks in multimodal analysis, it highlights how creators strategically integrate humour, stance, and visual/auditory stimuli to construct meaning and shape user engagement on social media.

2. Methodology

2.1. Materials

This study examines three *BuzzFeed's Tasty's* recipe videos selected among the files composing a larger corpus, the *som_cor*, the composition of which is shown in Table 1 (Battista 2025a). The *som_cor* considers Facebook, Instagram, and TikTok, as the most popular social media platforms at the moment of writing for three age cohorts: Facebook attracts people aged 25 to 34, Instagram users are typically 18 to 24, while TikTok is widely used by the youngest cohort. The *som_cor* considers two timespans: January 1st to January 7th, 2020 (the week before the first COVID-19 case in the USA), and May 11th to May 18th, 2023 (the week immediately following the end of the pandemic-related emergency state). Additionally, all posts met the following criteria: being in a video format; showing a recipe; featuring at least one *Tasty* creator; being an original *Tasty* production (as opposed to videos created by other brands and reshared by *Tasty*). Each video was assigned an identifier based on its social media platform, publication year, and progressive ID number; then, they were transcribed retaining fillers, self-repair, laughter, and any other pragmatically relevant feature.

	Source	Year	Tokens	Number of selected videos	Average duration of videos
	Facebook	2020	23,195	3	3:39 minutes
		2023		19	4:59 minutes
	Instagram	2020	3,140	0	-
		2023		7	50 seconds
	TikTok	2020	2,628	0	-
		2023		8	57 seconds
Total			28,963	37	

Table 1
Composition of the *som_cor* (Battista 2025a, p. 70).

As this explorative study intends to show how well-established frameworks can be adapted to the analysis of social media content, one video per platform was randomly selected by entering the video IDs into an online random choice generator (RandomJet, n.d.). The materials analysed for this article are: FB23_5 (*Tasty* 2023c); IG23_9 (*Tasty* 2023a); TT23_5 (*Tasty* 2023b). While limited, this exploratory corpus ensures genre consistency across three platforms, since all the videos share the same features (see the abovementioned selection criteria for the *som_cor*), facilitating comparative analysis. Moreover, considering multiple social media networks enables a cross-platform analysis (Brunner and Diemer 2022), which mirrors the brand's approach to content creation (Battista 2025b) and offers an insightful perspective in the analysis of online discourse.

2.2. Methods

This study relies on a multimodal discourse analysis approach bridging multimodality and pragmatics. In line with recent research emphasising the importance of considering multiple semiotic modes (Vásquez 2022), the aforementioned *BuzzFeed's Tasty's* videos have been treated as multimodal texts.

Each video has been annotated for its multimodal features using a dedicated digital tool, ELAN (Max Planck Institute for Psycholinguistics 2023), which allows researchers to create their own custom annotation tiers and to generate annotation plots. The annotations (Figure 1) considered camera shots (close-up, medium, long), vertical camera angles (high, eye-level, low) and creator/food screen time.

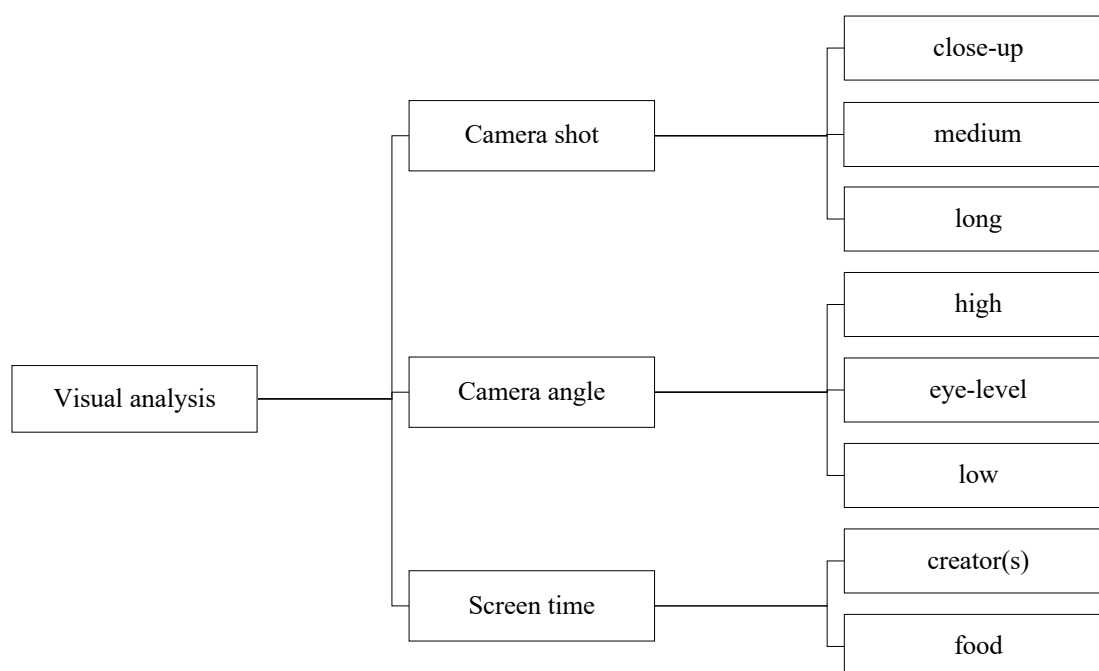


Figure 1
ELAN annotation tiers.

Creator/food screen time acts as a proxy for interactional focus across recipe stages (for instance, interpersonal function and procedural evidence). Camera shots and angles are recognised in multimodal discourse analysis as key tools to visually enact interpersonal relations and pragmatic meanings (Kress and van Leeuwen 2020; Machin and Mayr 2023). More specifically, shot distance indexes social distance and proximity and involvement, while vertical angles encode relative power or alignment between participants (viewers and creators, in this study). However, camera shots and angles may also depend on the specific function of the audiovisual product; in the context of a recipe video, these parameters (as well as creator/food screen time) may contribute to visually mediating sensory perceptions and may carry pragmatic values.

Consequently, the videos have been analysed manually for their pragmatic functions, namely for the illocutionary force (Austin 1962) of specific segments of the videos; stance markers, specifically adjectives, revealing the speakers' affective or evaluative stances (Kiesling *et al.* 2018; Kiesling 2022); and (im)politeness strategies, both linguistic and non-linguistic (Culpeper *et al.* 2017). Austin's speech act theory (1962) guided the analysis of the illocutionary force, which is the communicative intent behind utterances, thus distinguishing between directive instructions, assertive statements, and expressive evaluations. This made it possible to analyse the acts the speakers perform and allowed the researcher to better identify instructional, interactive, and engaging moments in the videos. For this part of the annotation, both linguistic and non-linguistic or contextual cues have been considered. In

addition to speech acts, this study explored the speakers' personal stances and positioning to express alignment or misalignment with content or interlocutors. Stance markers such as evaluative adjectives, intensifiers, hedges, and other verbal cues potentially disclosing affective or epistemic attitudes have been identified (Kiesling 2022). The videos have been annotated for evaluative words describing foodstuffs and cooking processes, markers of certainty and doubt during the preparation, and any kind of linguistic resource signalling social involvement. Non-verbal cues have also been noted, since stance can be conveyed multimodally. This type of analysis can help understand how speakers position themselves, and how they evaluate actions and ingredients. Politeness strategies describe how speakers save or threaten one's and others' faces through the use of friendly language, hedges, humour (Culpeper *et al.* 2017). Although impolite expressions are not necessarily offensive, they may also be used playfully to foster social affiliation (Culpeper *et al.* 2017). As for the previous frameworks, non-verbal signals, and gestures specifically, have been considered. This part of the analysis contributed to the understanding of how *Tasty* creates and maintains user engagement and social alignment in the presence of multiple speakers.

Since the materials were posted on three different platforms, a cross-platform approach has been adopted (Brunner and Diemer 2022). This involves comparing the annotated features across distinct social media platforms, possibly due to varying goals and intended audiences. As a matter of fact, the descriptive annotations provided by ELAN supported the qualitative insights, indicating the frequency and distribution of multimodal strategies across platforms, which allows the researcher to jointly consider multiple pragmatic and multimodal cues contributing to meaning-making processes in online discourses. However, these strategies have also been examined relying on speech act theory, stancetaking, and politeness strategies to determine how they function in context, in which way(s) they contribute to meaning-making processes, and with which aims. Taken together, these frameworks provide a multimodal pragmatic perspective on the interaction between speech acts, stance, politeness, and multimodality in the construction of the 'expert-friend' persona and user engagement across platforms.

3. Analysis

This section of the paper presents the analysis in three separate subsections, one for each video and social media platform, namely Facebook, Instagram, and TikTok. Then, the resulting insights will be commented and compared in the Discussion and Conclusion section of the article.

3.1. Facebook

This Facebook video shows how to prepare a delicious, homemade Mac ‘n’ Cheese. The visual annotation of the video demonstrates that visual deixis supports verbal instructions. More specifically, medium, eye-level shots dominate throughout the preparation phase, while close-ups are preferred when the focus is on textures or showing ingredients in greater detail; cooking-related sounds can occasionally be heard, such as sizzling, which makes the food particularly appealing. Nevertheless, this is the only sensory perception which users can access, as the focus is not on the sensory/sensual appeal of the dish, but on its status as a comfort food. A cross-comparison of the visual annotation with the rhetorical structure of the video shows a pattern. The creator mostly appears at the beginning and at the end of the video to initiate and close the interaction and is only shown during the preparation of the food when making jokes. This results in contributing to user engagement, while keeping interpersonal work as distinct from instructions.

The second step of the analysis involved a manual annotation of the transcript to identify speech acts. Assertives dominate, contributing to the narrative organisation of the video. Directives follow, as is predictable in a recipe video (e.g. “Crush up some crackers”); however, only half of them are procedural imperatives, while the others are inclusive acts used to introduce a new step (e.g. “We’re gonna make our topping”) or to make a funny remark (e.g. “Let’s get cheesy!” when adding a considerable amount of cheese). To conclude, some acts are used as engagement devices, namely questions, as in “This recipe is definitely not healthy, but you know. **Who’s keeping track these days?**” (emphasis added).

The third step of the analysis is the identification of stancetaking devices. In this transcript, evaluative language dominates and is used to describe the dish with sensory adjectives, such as *cheesy* or *silky*, alongside more general use words, for instance *perfect*, *ultimate*, and *beautiful*. These evaluatives are frequently accompanied by intensifiers, such as *so* and *extra*. They serve a double function: on the one hand, they provide useful cues for users who may wish to replicate the recipe at home; on the other hand, they provide accurate descriptions of the sensory qualities of the food which people watching the video would never otherwise be able to experience. Furthermore, evaluatives are relevant as in this video there are no explicit sensory cues (the only exception being the background sizzling sound).

This connection with the audience is further reinforced through the use of positive politeness, as shown by the extensive use of the pronoun *we*, the second person singular pronoun *you*, *let’s*, and the vocative *guys*. These linguistic devices preserve solidarity and create social affiliation, allowing users to make their own cooking choices while also asserting Alix’s expertise, as in “You can also use whatever kind of cheese you’d like. Again, **we’re not**

being picky here” (emphasis added). Multimodally, social affiliation is conveyed by the setting of the video, which is Alix’s own home kitchen and not a professional studio, as this video was originally shot during the COVID-19-related time of social isolation, which pragmatically indexes authenticity and relatability. Similarly, the creator often makes playful remarks and engages in self-mockery which fosters affiliation, as in the abovementioned rhetorical question about not really caring about eating healthy food, or in the following example: “So it’s not exactly ideal to use this metal whisk on my non-stick pan, but it’s all I got right now. Don’t come for me, Internet. Be gentle with me. I’m fragile right now. I’ve been alone so long” (it should be noted that Alix used to make fun of herself for not having been in a relationship for a long time in most of her videos). It is in these funny moments that multimodal cues contribute to accurate interpretations based on the creator’s gestures, facial expressions, and laughter. Finally, the concluding move “I hope it inspires you to do some good cooking of your own. If you want me to try out anything else, let me know, and I’ll see you guys later” maintains rapport and establishes continuity, invites user feedback, and explicitly signals Alix’s commitment to creating more videos, possibly in response to users’ suggestions.

Overall, the Facebook recipe video foregrounds clarity while relying on multimodality for sensual and sensory appeal to attract users. Alix positions herself as an expert and a down-to-earth, relatable, and fun person. Moreover, multimodality contributes to a pragmatic scaffolding of the video, marking interpersonal solidarity and technical instructions.

3.2. *Instagram*

The Instagram video considered in this analysis shows how to make a dupe version of the famous Crumbl Cookies. The longest section of the video is the preparation, which is followed by a presentation of the final result. The introduction shows the cookie for a second, just enough to break it in half and attract users. Then, a playful moment follows. It is organised in such a way as to evoke TV quiz shows: the creator holds a tiny microphone and presents two cookies on two tables, challenging users to identify the original Crumbl Cookie and its dupe and share their guesses in the comments. Additionally, the two cookies are shown in detail through a magnifying glass and are both positioned on a Crumbl Cookies packaging box, further reinforcing the visual comparison between the original product and the dupe. This is a popular and effective strategy to capture users’ attention and encourage engagement, as it motivates them to watch the whole video and participate actively, in line with the viral and participatory culture of social media content (Jenkins 2006). Furthermore, the two cookies are shown in detail using a magnifying glass and both positioned on a Crumbl Cookies packaging box.

The visual annotation of the video shows a dominant presence of medium shots at eye-level or from a high angle: the former highlights details, while the latter is reminiscent of the so-called ‘top-down’ videos. As a matter of fact, this video heavily relies on multimodality to efficiently provide instructions in a limited span of time (59 seconds).

The presence of on-screen ingredients, which function as visual deixis anchoring and compressing procedural information, allows the creator to save some time, taking advantage of the fact that social media videos can be rewatched by users at any time. Similarly, the last scene of the video shows a close-up of a cookie being broken in half with an on-screen caption reading “Get the full recipe on Tasty.co!”, which redirects users to the brand’s website for a written recipe, while allowing the creator to shoot a short and essential video.

Close-ups are used to show the most luxurious ingredients, such as chocolate chips, or the consistency of the dough. Therefore, camera shot choices convey sensory perceptions. Sensory perceptions are also expressed through language. This is done by using assertive speech acts, which account for about 83% of the transcript, and creative language as in “It’s time to fluffify” when using a stand mixer to mix all the ingredients; although the verb *fluffify* does not exist in standard English, it is easy to understand that it describes the final consistency that the dough should have. Thus, it acts as a controller, namely as additional instruction ensuring a recipe step has been successfully completed (Diemer 2013). In spite of this burst of creativity, the video does not make extensive use of evaluative language or playful remarks, which, when present, are wisely blended with the instructions, thus making the style direct and essential, providing accurate instructions while still being engaging. This may be due to the affordances of the means: Instagram videos are typically short, and this one only lasts 59 seconds, so it is paramount to effectively balance accurate procedural steps and user engagement strategies.

Another relevant category of speech acts is that of directives, which are typically softened, as in “You’re gonna scoop the cookies,” rather than realised as bare imperatives, such as “Mix it all together,” thereby exemplifying positive politeness strategies. Multimodally, they align with medium eye-level shots, which contribute to clarity. This and the use of the inclusive *we* and mitigated directives via *you’re gonna* + verb contribute to both positive politeness and social affiliation between the creator and the audience. Additionally, there are a lot of moments in which the creator tries to engage her audience, for instance by encouraging user feedback (“Let us know in the comments”).

A multimodal inspection of the video reveals that it is shot by the creator in what seems to be a private kitchen, which reinforces social affiliation with the audience. Moreover, food is allocated more screen time than the creator, who is foregrounded in specific moments, namely brief jokes (as in “It is the

next day and I've got a sweet tooth" to suggest the appealing look of the cookies after resting in the fridge overnight and before baking) and technical details (e.g. "Now I hold in my hands two room temperature eggs that we are going to add in, why? Because they mix better with the batter and they rise more easily"). The combination of accessible technical explanations, and playful and relatable remarks contributes to signalling expertise and friendliness at the same time.

To conclude, the Instagram recipe playfully merges a fun style with accurate technical instructions. The creator Krista positions herself as an expert and a fun friend, who uses a variety of strategies, such as politeness and jokes, to make the recipe accessible and build solidarity with her audience.

3.3. TikTok

Jasmine's video posted on *Tasty's* TikTok account features a special guest for a special occasion: the creator's mum for Mother's Day. The brief introduction foregrounds the affective and authentic nature of the recipe, as it presents Taiwanese cold noodles with sesame sauce as prepared by Jasmine's mother. A longer preparation phase follows and includes some digressions with heritage positioning (Jasmine has Taiwanese origins) and affective stance (due to the relationship between the two cooks). The video ends with a tasting moment on the same heritage-related and affective-focused notes characterising the content.

The visual annotation indicates a dominance of medium shots with some close-ups of the noodles or the sauce being poured. The creator and her guest are shown throughout the video as they cook in a home setting, which further reinforces a sense of relatability and authenticity.

The analysis of speech acts reveals that assertives dominate and are mostly used to describe the dish's appearance and sensory qualities, as in "The sauce is really nutty". Directives are used to provide instructions, but they are mostly softened, either by being presented as suggestions or through positive politeness strategies and inclusive forms such as "You want to make sure the noodles are cold" and "We're gonna toss the noodles." The video also makes frequent use of expressives, for example "Oh my gosh, that looks amazing," and rhetorical questions that promote audience engagement, as in "Who doesn't love cold noodles in summer?". Furthermore, there are some humorous exchanges between Jasmine and her mother, which contribute to authenticity, spontaneity, and relatability. For instance, because the video is shot to celebrate Mother's Day, Jasmine offers to shred the chicken to help her mother; however, the woman had already completed the task, which makes her daughter's help superfluous.

Stancetaking reveals that the creator's tone conveys expertise, relatability, and affection at the same time. Evaluative adjectives are not

particularly frequent and are used either to provide useful procedural information or to describe the food at the end. The main stance is related to Jasmine's identity: the recipe becomes a way for her to connect with her origins and with her mother, while also showcasing relatable expertise. In this way, she connects with her audience through parasocial interaction (Horton and Wohl 1956).

In spite of being relatively short (1:14 minutes), this video relies on multimodality to add a layer of complexity to the recipe. Firstly, it has already been mentioned that the setting is Jasmine's private kitchen, which contributes to solidarity and 'relatable coolness'. Secondly, the camera is positioned on the counter in front of a colourful portable stove with a *Tasty* pot, which is a smart marketing move, suggesting that creators use the brand's products on a daily basis without necessarily promoting them. Thirdly, the full recipe can be accessed in the video description, and on-screen captions are only used at the end of the video, during the food tasting, when Jasmine and her mother speak Taiwanese, which further reinforces the heritage and identity positioning of the creator and her guest.

Jasmine's recipe can be described as a typical social media video, providing instructions, while making use of the affordances of the means to provide accurate, rich, but not tedious content. Indeed, as the creators only mention the steps of the recipe (as the automated subtitles also show), leaving more detailed information in the video description, for instance regarding the needed quantity for each ingredient. Furthermore, multimodality contributes to the accessibility of video by using on-screen captions to translate Taiwanese into English, so that the spontaneity and authenticity of the interaction between the creator and her mother are preserved.

4. Discussion

The analysis section adopted a descriptive approach to present the three videos randomly selected for this study from a wider corpus of *Tasty* videos. As previously mentioned, since the videos were posted on three different social media networks, a cross-platform approach (Brunner and Diemer 2022) will be integrated to highlight similarities and differences.

4.1. 'Expert-friend' persona

BuzzFeed's Tasty's videos have a distinctive feature regardless of the posting platform since they blend instructional content with entertainment, thus fully aligning with *infotainment* content, which has gained popularity on the Internet (YEC, n.d.). This emerges from the videos analysed here: the creators instruct viewers on how to cook a dish, while engaging and entertaining them. All three

videos construct an approachable “expert friend” persona, pragmatically and multimodally negotiated through a balance between expertise and relatability, which makes the content accessible and, at least seemingly, authentic (Battista 2025a). These findings illustrate how the methodological frameworks used for the analysis intersect and operate as a unified set of layers. For instance, stance is conveyed both linguistically and visually; politeness is expressed via verbal humour and camera shots; speech acts gain force through multimodality. This extends existing models of influencer authority (Abidin 2018) by showing the balance between expertise and relatability in specialised multimodal digital contexts. Nonetheless, the tone, style, and interactional features vary across Facebook, Instagram, and TikTok, reflecting previous research on the topic (Haenlein *et al.* 2020).

Alix, on Facebook, adopts a narrative and empathetic tone. She warmly welcomes the audience into her apartment – albeit only virtually – and sympathises with them, as everyone at that time was experiencing social isolation. This situates the video as a source of entertainment during the difficult period of the pandemic, when social media was one of the few, if not the only, forms of escape from the real world, and fosters an emotional connection with viewers. Furthermore, Alix makes several jokes, even relying on self-mockery, which makes her appear entertaining, approachable, and human (Matwick and Matwick 2019). This is possible due to the generally longer length of Facebook videos as opposed to other platforms. This abundance of humour, together with the use of informal language, serves as a positive politeness strategy, making the audience feel like close friends rather than an anonymous mass to instruct (Jones 2024). This can foster trust, familiarity, and can position the host as an authoritative cook and as a relatable peer at the same time, thus epitomising the expert/non-expert balance which is a key feature in food blogging (Cesiri 2016).

On Instagram the tone is upbeat and concise. The entertaining side of the video is achieved through a playful hook presenting the content as a challenge for users to guess which is the original and which is the dupe Crumbl Cookie. While this move encourages user participation, this is also a ‘fake’ challenge, since its real purpose is to demonstrate that there are no visible differences between the two cookies, thereby enhancing the perceived appeal and quality of the recipe. Additionally, this situates the video within the trend-driven social media culture, where content often revolves around viral challenges or ‘hacks’ (Barros *et al.* 2025; Biolo 2023). The language is friendly and colloquial, but focused on action, visual instructions, and visual appeal rather than on personal narratives or digressions. As a matter of fact, the creator highlights the originality of the recipe, which makes a popular product accessible to anyone who is willing to cook, and conveys relatability and fun through Krista’s communicative and creative style. As a result, the ‘expert’ persona here is more

of a relatable presenter/entertainer rather than a close friend due to the lack of intimate, heart-to-heart exchange.

The TikTok video exemplifies authenticity and personal storytelling on social media, reflecting TikTok creators' tendency to adopt a casual and relatable style and to share real, and even vulnerable, moments (Barta and Andalibi 2021). Although the case study of this article is a brand and not a personal profile, some of the *Tasty* creators stand out for personal storytelling. This has already emerged in Alix's video, as she usually mocks herself for being single, but it is even more evident in Jasmine's content, who often invites her personal acquaintances in her videos and typically cooks Taiwanese(-inspired) dishes. In this specific video, the creator is with her mother for Mother's Day at her own apartment, which immediately creates a warm family atmosphere. The tone is intimate, the emotional bond is central, and it is clear that some of the dialogue between the two is spontaneous, for example when Jasmine offers to help shred the chicken – which her mum had already finished doing – and when her mother shyly reacts to Jasmine saying, “I love you so much!”. These genuine interactions and Jasmine's typically relatable style situate this video within the ‘slice of life’ trend and foster authenticity (Barta and Andalibi 2021; Yau and Reich 2019). Consequently, the creator positions herself as the prototypical ‘expert-friend’, introducing her own family to viewers, showing her pride for her origins, as well as enthusiasm for quality family time and, of course, cooking (Zeng *et al.* 2021), thus associating expertise and relational authenticity, creating both affective and affiliative bonds (Kiesling 2022).

Overall, it can be concluded that the three videos considered for this study aim at authenticity, aligning with social media strategies for user engagement and, consequently, monetisation (Cascio Rizzo *et al.* 2023; Nunes *et al.* 2021). However, different strategies are adopted: Facebook emphasises personal narratives, solidarity, and humour; Instagram focuses on trendy and fun engagement; TikTok prioritises personal storytelling and realness. This proves that the ‘expert-friend’ is both a successful and an adaptable strategy.

4.2. *Speech acts and stance*

The annotation of speech acts confirmed a dominant pattern, namely the dominance of assertives and directives across the dataset. While this is not surprising due to the instructional nature of the videos (Lacsina 2023), it is interesting to observe how their form and pragmatic force vary across platforms.

A recurring pattern is that of softening directions. On Facebook, this is achieved through the use of inclusive pronouns and narrative, personal framing. On Instagram, directives are concise and paired with visual support, which is an efficient use of multimodal resources to keep the video essential

and short. On TikTok, directions are presented as suggestions, which aligns with Jasmine's affiliative style and with previous pragmatic research (Shaarawy 2015) suggesting that the nature of imperatives depends on agent type. In this case, the softened use of directives resonates with the "expert friend" persona.

Assertives are equally relevant and variable: mostly descriptive on Facebook, procedural on Instagram, and affective or evaluative on TikTok. This pattern reflects Kiesling's (2022) insight that stance is a layered process, since epistemic stance prevails on Facebook, procedural stance on Instagram, and affective stance on TikTok.

Another important finding concerns politeness. Although impoliteness can sometimes be used playfully to foster social affiliation, this was not observed in the videos analysed. Instead, all three videos rely on positive politeness to reinforce solidarity through inclusive we pronouns, hedged directions, humour, and personal storytelling. This may suggest that affiliation, together with the creation and maintenance of a safe environment, is a priority for a commercial brand like *Tasty* (Creelman 2022; Vandevyvere 2016).

4.3. Engagement

A common feature that these videos display is inviting audience engagement. Instagram is the most explicit as it directly asks users to make their guess about the cookies in the comments. This strategy fosters a participatory experience revolving around the core of the video (baking a Crumbl Cookie dupe) (Pletikosa Cvijikj and Michahelles 2013), which is crucial on a platform where visibility may depend on engagement. The Facebook video invites users to comment but only at the end, when they are encouraged to suggest potential future recipes. This is useful for creators to know what their audience might appreciate, and engages viewers directly in the content creation process, giving the impression that the creator will be listening to them, thus fostering loyalty (Ertz *et al.* 2024; Malodia *et al.* 2024). In any case, posing a question to the audience and encouraging interaction in the comment section is typically used on social media for involvement (Lacsina 2023).

Inviting comments is not the only way to encourage user participation. All the videos analysed use inclusive we, which creates a shared cooking experience and frames it as a collaborative activity between the host and the viewers, making them feel involved. This involvement is further reinforced multimodally through close-up shots that show the dish or ingredients in detail. Such shots serve two purposes simultaneously: creating an immersive experience for the audience and supporting the cooking instructions, which is one of the most typical and ancient uses of images in recipes (Notaker 2017). Furthermore, parasocial interactions are present (Giles 2002), as all the creators 'imagine' the actual presence of their audience. On TikTok, viewers are

positioned as learning an authentic Taiwanese recipe and as witnesses of an equally authentic family interaction, whereas on Instagram users are addressed as the contestants on a TV quiz show. However, Facebook presents the most interesting case. Alix directly addresses the audience by sharing personal anecdotes and by mocking herself, for instance when she uses a metal whisk on a non-stick pan. This results in anticipating potential critical commentaries and defusing negativity while acknowledging users in a humorous way (Culpeper *et al.* 2017), strengthening social connection and demonstrating pragmatic awareness in a social context by using humour and self-mockery to mitigate potentially face-threatening acts (as well as acts potentially undermining her cooking expertise) (Sundar and Cao 2020).

Overall, these examples demonstrate that engagement is not only a social media marketing strategy, but also functions as a pragmatic act of affiliation, performing illocutionary work by including the audience as a co-participant and fostering solidarity. Therefore, engagement is a speech act co-constructing the social dimension of digital food discourse, fostering loyalty, and supporting the brand's communicative goals (Abidin 2018; Ertz *et al.* 2024).

4.4. Multimodal strategies

The multimodal organisation of the videos and the orchestration of visuals, sounds, gestures, and words play a central role in meaning-making processes. Although it can be argued that all communication is inherently multimodal (Kress and van Leeuwen 2020), digital contexts offer complex interplays of semiotic modes.

All the videos analysed rely on visual deixis, thus using visual resources to support or complement verbal instructions, for instance by showing a full list of ingredients or additional information as on-screen captions. TikTok is the only exception, as on-screen elements are only used to make the video fully accessible to people who do not speak Taiwanese.

The use of camera angles and shots varies across social media platforms, reflecting different multimodal and pragmatic priorities. On Facebook and TikTok, medium eye-level shots dominate, and the creator has the same screen time as the food, thus reflecting the balance between showing a recipe and offering personal narratives. Instagram provides the most close-ups, echoing the original *Tasty* top-down videos, whose priority is that of showing how to make a recipe with minimal social and personal engagement. These camera choices are therefore not merely aesthetic but contribute to pragmatic meaning-making. Close-ups highlight textures or results and typically accompany procedural and evaluative acts, visually reinforcing the illocutionary force of the instruction and/or its outcome. Medium eye-level shots of the creator usually coincide with softened directives, assertives, and affiliative talk, as the equal (eye-level) perspective visually enacts proximity and is thus consistent

with the 'expert friend'. High or top-down angles align with step-sequencing and visual clarity, effectively guiding the user through instructions. Consequently, camera work is pragmatic in that it extends the illocutionary work of speech acts into the visual mode.

Soundscapes diverge. The Facebook video is the only one with auditory food-related sensory perceptions. The Instagram video has a very quick pace and only the voice of the creator is heard. Jasmine's TikTok video prioritises a realistic soundscape, with clinking bowls and laughter, thus multimodally enhancing authenticity. These differences confirm that the illocutionary force of a speech act can extend beyond language to multimodal communication (O'Halloran 2011; Payrató 2017) and multisensory experiences (Keating and Łapińska 2023).

Platform affordances affect content creation, even though they do not depend on brand's or creators' choices. Facebook typically has a higher tolerance for longer content, which enables more digressions, as opposed to Instagram, whereas TikTok's algorithm favours emotional and personal bonds. This exemplifies that recipes as a digital genre mediate form, action, and medium to foster novel, multimodal, participatory environments (Miller 2015).

These insights confirm the centrality of multimodality in social media discourse as a tool to expand the illocutionary force of speech acts (Austin 1962), since directives gain clarity and expressives are intensified. Consequently, this study suggests that speech acts and visual communication should not be treated as separate layers, but as complementary meaning-making tools (Payrató 2017), thereby amplifying their force for instructive, humorous, and affiliative goals (Zappavigna 2015).

5. Conclusion

The present study considered *BuzzFeed's Tasty* as a case study to explore how cooking videos on social media employ multimodal strategies for instruction, affiliation, humour. Inspired by pragmatics and relying on the frameworks of multimodality (Kress and van Leeuwen 2020), politeness (Culpeper *et al.* 2017), speech acts (Austin 1962), and stancetaking (Kiesling 2022), this article shows the adaptability of pragmatic frameworks to the analysis of multimodal digital genres and notes similarities and differences across three social media networks through a cross-platform approach (Brunner and Diemer 2022).

Across all platforms, pragmatic functions, mainly instruction and affiliation, are realised multimodally through the interplay of speech acts, stance markers, politeness, camera work, and humour. Nevertheless, each platform foregrounds a different communicative style: Facebook prioritises a colloquial and friendly tone, Instagram a fun and concise style, and TikTok spontaneity. Overall, two common patterns emerge: the creators' positioning

as “expert friends” and their pursuit of authenticity as strategies to boost user engagement.

This study entails some limitations. Firstly, the dataset is highly specialised, since only three random videos – one per platform – have been considered. Although a larger corpus was available, the main goal of this study was to rely on multiple frameworks to explore their interplay and effects on social media content. Since the annotation process was entirely manual and subject to time and resource constraints, only a limited number of videos could be analysed. Secondly, only one brand and its videos posted in May 2023 have been analysed. Thirdly, the annotation was carried out by a single researcher. All these limitations point to a shared gap which could be addressed in future research: expanding the dataset and, ideally, involving more annotators could make the analysis more representative since, at this stage, it is not possible to make generalisations. Instead, emerging trends may provide food for thought for future studies. Finally, the investigation could benefit from computational tools for a quantitative perspective on the issue.

Nonetheless, the study contributes to pragmatics by demonstrating how traditional concepts can be adapted to new platforms while extending verbal communication to multimodal resources to co-construct pragmatic functions. A multimodal pragmatic perspective reveals that instruction and affiliation are realised through language, camera work, editing, sound, humour, and stance markers. This article also has implications for genre theory, as it suggests that a historically stable genre such as the recipe still offers room for adaptation to the participatory culture of social media without losing its instructional nature. Bridging pragmatics and multimodality can provide valuable tools for understanding the complexities of digital communication. Furthermore, while it does not offer definitive conclusions, this study opens up new avenues for research on digital food genres. Ultimately, a recipe is much more than a set of ingredients and procedural steps: it can become a site of stancetaking, social affiliation, and sensory perception.

Bionote: Alessia Battista is a Researcher in English Language, Translation, and Linguistics at Tor Vergata University of Rome. She holds a PhD (*Doctor Europaeus*) in “Euro(pean) Languages and Specialised Terminologies” from Parthenope University of Naples (Italy) in *cotutelle* with the University of Artois (France). She was a Research Fellow at the University of Salerno and has previously taught at Suor Orsola Benincasa University of Naples, the University of Salerno, and the University of Naples Federico II. She was awarded the Emerging Scholar Award granted by the American Name Society. Her latest monograph, *Oltre il binarismo di genere. Analisi dell'identità butch*, has been published as an open-access book by Suor Orsola University Press. Her research interests focus on English sociolinguistics, ESP, (social) media discourse, multimodality, corpus linguistics, critical discourse analysis, and AI.

Author's address: alessia.battista@uniroma2.eu



References

- Abidin C. 2018, *Internet Celebrity: Understanding Fame Online*, Emerald Publishing, Bingley.
- Adami E. and Jewitt C. 2016, *Special Issue: Social Media and the Visual*, in "Visual Communication" 15 [3], pp. 263-270.
- Andersson M. 2024, *E-Mpoliteness – Creative Impoliteness as an Expression of Digital Social Capital*, in "Journal of Politeness Research" 20 [2], pp. 227-248.
- Arendholz J., Bublitx W., Kirner M. and Zimmermann I. 2013, *Food for Thought – or, What's (in) a Recipe?: A Diachronic Analysis of Cooking Instructions*, in Gerhardt C., Frobenius M. and Ley S. (ed.), *Culture and Language Use*, John Benjamins, Amsterdam, pp. 119-138.
- Ashley C. and Tuten T. 2015, *Creative Strategies in Social Media Marketing: An Exploratory Study of Branded Social Content and Consumer Engagement*, in "Psychology & Marketing" 32 [1], pp. 15-27.
- Austin J.L. 1962, *How to Do Things with Words: The William James Lectures Delivered at Harvard University in 1955*, Urmson J.O. and Sbrisà M. (ed.), Oxford University Press, Oxford.
- Baldauf H., Develotte C. and Ollagnier-Beldame M. 2017, *The Effects of Social Media on the Dynamics of Identity: Discourse, Interaction and Digital Traces*, in "ALSIC. Apprentissage des Langues et Systèmes d'Information et de Communication" 20 [1].
- Barros P.A.A., de Paula Matos J., Bittencourt Rodrigues M., da Costa R.J. and Martins Horta P. 2025, *Food Digital Marketing on Social Media: Trends and Strategies of Brazil's Leading Meal Delivery App*, in "Frontiers in Nutrition" 12, no. 1620348.
- Barta K. and Andalibi N. 2021, *Constructing Authenticity on TikTok: Social Norms and Social Support on the "Fun" Platform*, in "Proceedings of the ACM on Human-Computer Interaction" 430, pp. 1-29.
- Battista A. 2025a, *Decoding the Language of Food: A Corpus-Assisted Multimodal Analysis of BuzzFeed's Tasty's Social Media Content*, PhD dissertation, 'Parthenope' University of Naples/University of Arras, Naples/Arras.
- Battista A. 2025b, *Inspiring People and Migrating to New Social Media Platforms: A Corpus-Based Analysis of BuzzFeed's Tasty's Evolution over the Years*, in Stoian C. (ed.), *Language, Culture and Business*, Cambridge Scholars Publishing, Newcastle upon Tyne, pp. 135-174.
- Biolo F. 2023, *TikTok and Gastronomy: A New Horizon on Digital Communication*, BA dissertation, University of Padova, Padua. https://thesis.unipd.it/retrieve/0301ae0c-4c0f-49b5-99a1-f4d5779e15b5/Francesco_Biolo_2013831_TESI%20%281%29.pdf (16.08.2025).
- Bhatia V. 2008, *Towards Critical Genre Analysis*, in Jones R. and Flowerdew J. (ed.), *Advances in Discourse Studies*, Routledge, London, pp. 166-177.
- Bouvier G. and Machin D. 2018, *Critical Discourse Analysis and the Challenges and Opportunities of Social Media*, in "Review of Communication" 18 [3], pp. 178-192.
- Brunner M.L. and Diemer S. 2022, *Cross-Platform Analysis*, in Vázquez C. (ed.), *Research Methods for Digital Discourse Analysis*, Bloomsbury, London, pp. 197-218.
- Cascio Rizzo G.L., Berger J., De Angelis M. and Pozharliev R. 2023, *How Sensory Language Shapes Influencer's Impact*, in "Journal of Consumer Research" 50 [4], pp. 810-825.
- Cesiri D. 2016, *The Blog Is Served: Crossing the "Expert/Non-Expert" Border in a Corpus of Food Blogs*, in "Colloquium: New Philologies" 1 [1], pp. 47-62.

- Cesiri D. 2020, *The Discourse of Food Blogs: Multidisciplinary Perspectives*, Routledge, London.
- Creelman V. 2022, “Thank You for Reaching Out.” *Brand Relationship Management and the Conversational Human Voice of Customer Care in Online Service Encounters*, in “Discourse, Context & Media” 46, no. 100572.
- Culpeper J., Haugh M. and Kádár D.Z. (ed.) 2017, *The Palgrave Handbook of Linguistic (Im)Politeness*, Palgrave Macmillan, London.
- Culpeper J., Haugh M. and Sinkeviciute V. 2017, *(Im)Politeness and Mixed Messages*, in Culpeper J., Haugh M. and Kádár D.Z. (ed.), *The Palgrave Handbook of Linguistic (Im)Politeness*, Palgrave Macmillan, London, pp. 323-355.
- Cunningham S. and Craig D. (ed.) 2021, *Creator Culture: An Introduction to Global Social Media Entertainment*, New York University Press, New York.
- Del Campo M.A. and Kehle T.J., 2016, *Autonomous Sensory Meridian Response (ASMR) and Frisson: Mindfully Induced Sensory Phenomena that Promote Happiness*, in “International Journal of School & Educational Psychology” 4 [2], pp. 99-105.
- Diemer S. 2013, *Recipes and Food Discourse in English – a Historical Menu*, in Gerhardt C., Frobenius M. and Ley S. (ed.), *Culture and Language Use*, John Benjamins, Amsterdam, pp. 139-156.
- Ertz M., Cao X. and Barragán Maravilla J.M. 2024, The Prosumer, in “Encyclopedia” 4 [3], pp. 1263-1278.
- Garzone G. 2017, *Il genere ricetta in lingua inglese. Dalla tradizione ai food blog tra stabilità ed evoluzione*, in “LCM – La Collana / The Series”, pp. 37-56.
- Giles D.C. 2002, *Parasocial Interaction: A Review of the Literature and a Model for Future Research*, in “Media Psychology” 4 [3], pp. 279-305.
- Görlach M. 1992, *Text-Types and Language History: The Cookery Recipe*, in Rissanen M., Ihalainen O., Nevalainen T. and Taavitsainen I. (ed.), *History of Englishes*, De Gruyter Mouton, Berlin, pp. 736-761.
- Grzenkiewicz M. and Wildfeuer J. 2025, *Addressing TikTok’s Multimodal Complexity: A Multi-Level Annotation Scheme for the Audio-Visual Design of Short Video Content*, in “Digital Scholarship in the Humanities” 40 [4], pp. 1143-1166.
- Haenlein M., Anadol E., Farnsworth T., Hugo H., Hunichen J. and Welte D. 2020, *Navigating the New Era of Influencer Marketing: How to Be Successful on Instagram, TikTok, & Co.*, in “California Management Review” 63 [1], pp. 5-25.
- Horton D. and Wohl R.R. 1956, *Mass Communication and Para-Social Interaction: Observations on Intimacy at a Distance*, in “Psychiatry” 19 [3], pp. 215-229.
- Jenkins H. 2006, *Convergence Culture: Where Old and New Media Collide*, New York University Press, New York.
- Jones R.H. 2024, *Discourse Analysis: A Resource Book for Students*, Routledge, London.
- Jucker A.H. 2024, *Speech Acts: Discursive, Multimodal, Diachronic*, Cambridge University Press, Cambridge.
- Kang J., Hong S. and Hubbard G.T. 2020, *The Role of Storytelling in Advertising: Consumer Emotion, Narrative Engagement Level, and Word-of-Mouth Intention*, in “Journal of Consumer Behaviour” 19 [1], pp. 47-56.
- Keating M. and Łapińska J. 2023, *From Cookbooks to ASMR: Significance of Sound and Hearing in Culinary Recipes*, in “The Senses and Society” 18 [3], pp. 317-328.
- Kiesling S.F. 2022, *Stance and Stancetaking*, in “Annual Review of Linguistics” 8 [1], pp. 409-426.
- Kiesling S.F., Pavalanathan U., Fitzpatrick J., Han X. and Eisenstein J. 2018, *Interactional Stancetaking in Online Forums*, in “Computational Linguistics” 44 [4], pp. 683-718.
- Kress G.R. and van Leeuwen T. 2001, *Multimodal Discourse: The Modes and Media of*

- Contemporary Communication*, Arnold, London.
- Kress G.R. and van Leeuwen T. 2020, *Reading Images: The Grammar of Visual Design*, Routledge, London.
- Lacsina N. 2023, *Unveiling the Art of Food Vlogging: A Multimodal Discourse Analysis of Food Review Vlogs*, in "International Journal of Linguistics and Translation Studies" 4 [2], pp. 11-25.
- Machin, D. and Mayr A. 2023, *How to Do Critical Discourse Analysis: A Multimodal Introduction*, SAGE, Thousand Oaks (CA).
- Malodia S., Filieri R., Otterbring T. and Dhir A. 2024, *Unlocking Social Media Success: How Prosumers Drive Brand Engagement through Authentic Conversations with Consumers*, in "British Journal of Management" 35 [4], pp. 2197-2212.
- Matwick K. and Matwick K. 2019, *Storytelling on Cooking Shows*, in Matwick K. and Matwick K. (ed.), *Food Discourse of Celebrity Chefs of Food Network*, Springer, Cham, pp. 49-78.
- Max Planck Institute for Psycholinguistics 2023, *ELAN*, v. 6.7, Nijmegen. <https://archive.mpi.nl/tla/elan> (31.07.2025).
- McDonnell E.M. 2016, *Food Porn: The Conspicuous Consumption of Food in the Age of Digital Reproduction*, in Bradley P. (ed.), *Food, Media and Contemporary Culture: The Edible Image*, Palgrave Macmillan, London, pp. 239-265.
- Miller C.R. 2015, *Genre as Social Action (1984), Revisited 30 Years Later (2014)*, in "Letras & Letras" 31 [3], p. 3.
- Notaker H. 2017, *A History of Cookbooks: From Kitchen to Page over Seven Centuries*, University of California Press, Berkeley (CA).
- Notaker H. 2024, *The History of Cookbooks*, in *Oxford Research Encyclopedia of Food Studies*. https://oxfordre.com/foodstudies/view/10.1093/acrefore/9780197762530.001.0001/a_crefore-9780197762530-e-16 (19.08.2025).
- Nunes J.C., Ordanini A. and Giambastiani G. 2021, *The Concept of Authenticity: What It Means to Consumers*, in "Journal of Marketing" 85 [4], pp. 1-20.
- O'Halloran K.L. 2011, *Multimodal Discourse Analysis*, in Hyland K. and Paltridge B. (ed.), *Continuum Companion to Discourse Analysis*, Continuum, London, pp. 120-137.
- O'Halloran K.L., Tan S. and Lim-Fei M.K. 2014, *Multimodal Pragmatics*, in Schneider K.P. and Barron A. (ed.), *Pragmatics of Discourse*, De Gruyter, Berlin, pp. 239-268.
- Payrató L. 2017, *Pragmatics and Multimodality. A Reflection on Multimodal Pragmatics*, in Giora R. and Haugh M. (ed.), *Doing Pragmatics Interculturally*, De Gruyter, Berlin, pp. 293-312.
- Pérez-Llantada C. 2024, *Approaching Digital Genre Composing through Reflective Pedagogical Praxis*, in "Journal of English for Academic Purposes" 68 no. 101349.
- Pletikosa C.I. and Michahelles F. 2013, *Online Engagement Factors on Facebook Brand Pages*, in "Social Network Analysis and Mining" 3 [4], pp. 843-861.
- RandomJet. n.d., *Random Choice Generator*. <https://randomchoicgenerator.com> (19.08.2025).
- Sanfilippo M. 2023, *Ricette social. Forme del fare-culinario fra Instagram e TikTok*, in "E|C" 38, p. 314-334.
- Schlüssel H. and Frosh P. 2023, *The Taste of Video: Facebook Videos as Multi-Sensory Experiences*, in "Convergence: The International Journal of Research into New Media Technologies" 29 [4], pp. 980-996.
- Shaarawy H. 2015, *Imperative in Written English Cooking Recipes: A Cognitive Linguistic Analysis*, in "English Language and Literature Studies" 10 [1], pp. 125-168.
- Sundar A. and Cao E.S. 2020, *Punishing Politeness: The Role of Language in Promoting Brand Trust*, in "Journal of Business Ethics" 164 [1], pp. 39-60.

- Tasty 2023a, *You gotta try these Copycat Chocolate Chip Crumbl Cookies from @callmekristatorres. ICONIC!*, Instagram, 12 May 2023. <https://www.instagram.com/buzzfeedtasty/reel/CsKHqG2OE2L> (19.08.2025).
- Tasty 2023b, *Taiwanese Cold Noodle with Sesame Sauce*, TikTok, 14 May 2023. <https://www.tiktok.com/@tasty/video/7233083769475927339> (19.08.2025).
- Tasty 2023c, *Ultimate Mac "N" Cheese At Home With Alix*, Facebook, 14 May 2023. <https://www.facebook.com/watch/?v=559968569619718> (19.08.2025).
- Turkle S. 2015, *Reclaiming Conversation: The Power of Talk in a Digital Age*, Penguin Press, New York.
- Vandevyvere K. 2016, *Politeness Strategies in B2C Communication on Social Media*, MA dissertation, University of Gent, Ghent.
- Vásquez C. (ed.) 2022, *Research Methods for Digital Discourse Analysis*, Bloomsbury, London.
- Vidgen H. 2016, *Food Literacy*, Routledge, London.
- Wang Y. 2021, *Multimodal Analysis: Researching Short-Form Videos and the Theatrical Practices*, in "Proceedings of the 104th Association for Education in Journalism and Mass Communication". <http://hdl.handle.net/10125/76003> (19.08.2025).
- Yau J.C. and Reich S.M. 2019, *It's Just a Lot of Work: Adolescents' Self-Presentation Norms and Practices on Facebook and Instagram*, in "Journal of Research on Adolescence" 29 [1], pp. 196-209.
- YEC, Council Post, 2021, *The Development of Infotainment in Video Marketing*, in "Forbes". <https://www.forbes.com/sites/theyec/2021/01/13/the-development-of-infotainment-in-video-marketing> (1.08.2025).
- Zappavigna M. 2015, *Searchable Talk: The Linguistic Functions of Hashtags*, in "Social Semiotics" 25 [3], pp. 274-291.
- Zeng J., Abidin C. and Schäfer M. 2021, *Research Perspectives on TikTok and Its Legacy Apps: Introduction*, in "International Journal of Communication" 15, pp. 3161-3172.
- Zhang Y., Lucas M. and Pedro L. 2022, *The Design of Multimodal Short Videos: The Case of English Teaching Videos in TikTok*, in "INTED2022 Proceedings", pp. 5691-5699.

EMOJI IN R/HIKIKOMORI

A Corpus-assisted Study of Digital Politeness, Humour, and Identity

SILVIA CAVALIERI¹, SARA CORRIZZATO²,
VALERIA FRANCESCHI²

¹UNIVERSITY OF MODENA AND REGGIO EMILIA, ²UNIVERSITY OF VERONA

Abstract – In English-mediated digital communication, emojis play a crucial role in shaping pragmatic strategies, influencing expressions of politeness, humour, or persuasion. This study examines their functions as meaning-making tools in a two-million-word corpus from posts on the r/hikikomori subreddit, an online community of users who withdraw from social life and remain confined to their homes for long periods. Within this context, where norms differ from mainstream platforms, emojis can act as affective amplifiers, softening harsh statements or signalling empathy in discussions about isolation and mental health. Their use in self-directed humour reflects awareness of one’s condition and reinforces in-group belonging. Drawing on multimodal discourse analysis and pragmatic frameworks, and combining corpus-assisted methods with qualitative examination, we investigate the interplay between the 10 most frequent emojis and text at both macro and micro levels. Our findings show that the emojis under investigation function primarily as expressions of emotions and attitudes as well as emotional and pragmatic intensity enhancers; they are also employed as politeness markers mitigating face-threatening acts or as cues for humour and irony clarifying or subverting textual meaning. Emoji interpretation is highly context-dependent, shaped by linguistic and extralinguistic factors, and contributes to the realisation of speech acts by strengthening or softening communicative intentions. This study highlights the evolving pragmatic role of emojis as flexible, context-sensitive elements in digital discourse. It contributes to understanding multimodal communication in niche online communities and calls for further research into how emoji use varies across subcultures and platforms, shaping online identity and interaction.

Keywords: hikikomori; Reddit; emoji; online communities; cyberpragmatics.

1. Introduction¹

First introduced by psychiatrist Tamaki Saito in the late 1990s, the term “hikikomori” derives from two Japanese words: *hiku* (引), meaning “to

¹ Although the research was carried out jointly by the three authors, Silvia Cavaliere wrote Sections 3, 4, 5.1 and 5.2.1, Sara Corrizzato wrote sections 1, 5.2.3, 5.2.5, 6, and Valeria Franceschi wrote Sections 2, 5.2.2. and 5.2.4.

pull back” or “to withdraw”, and *komoru* (こもる), meaning “to seclude oneself” or “to shut oneself away”.

The term “hikikomori” was coined to refer to a prolonged, marked pattern of social withdrawal in which a person remains at home most days, avoids school or work, and restricts in-person social contact for at least six months, with associated functional impairment (Kato *et al.* 2019; Saito 1998). This behaviour was first formally categorised in Japan in the late twentieth century, but more recently the phenomenon has been observed worldwide, leading experts around the world to study its possible causes/consequences and identify new assessment tools (Rooksby *et al.* 2020).

Recent international reviews point out that hikikomori is not merely “shyness” or ordinary solitude but a sustained retreat from face-to-face participation in society. A range of factors have been implicated in the development of hikikomori, including dysfunctional parent–child relationships, intense social expectations, academic/working pressures, mental health problems, or various combinations of these components (Krieg and Dickie 2013; Neoh *et al.* 2023). Expert debate remains active regarding whether hikikomori should be classified primarily as a psychiatric disorder or as an adaptive response to socio-cultural constraints. Within this discourse, researchers frequently identify two sub-classes of individuals: *primary hikikomori*, characterized by social isolation which happens in the absence of a clearly identifiable external trigger, and *secondary hikikomori*, in which social withdrawal is understood as a secondary manifestation of an underlying condition such as a psychiatric illness, traumatic experience, or severe psychosocial problems (Suwa and Suzuki 2013; Li and Wong 2015).

Most cases still find a predominance of adolescents and young adults who report a progressive pattern in which the aforementioned parameters can be detected: initial school/academic refusal or work disengagement, inversion of sleep-wake cycles and narrowing of daily activities to the private setting.

Prolonged time spent at home is associated with increasingly intensive Internet use, e.g., gaming, streaming and social media, which often becomes the hikikomori’s central occupation (Li and Wong 2015; Tateno *et al.* 2019). Hence, compared to earlier decades, hikikomori today make far greater use of the Internet (a trend that reflects the role digitalisation has in everyday society). As online activities permeate contemporary society, it has in some days naturally become even more important for individuals who spend most of their time physically isolated. For hikikomori, the digital sphere seems to be the primary source for information, entertainment and social interaction (Krieg and Dickie 2013; Neoh *et al.* 2023; Suwa and Suzuki 2013).

Acknowledging the massive presence of this community in online spaces, recent studies have focused on hikikomori’s digital interaction with the aim to understand the role virtual communication can have in their life. The literature reveals a dual nature of online interaction: on the one hand, it

can have a restorative, even salvific power, while on the other, it has the potential to exacerbate pre-existing problematic behaviours. Kato *et al.* (2019) suggest that online platforms can serve as a compensatory social space for hikikomori, offering effortless opportunities for interacting with other people; therefore, Internet fora and social media platforms can help hikikomori get and keep in contact with other people, contributing to reducing total isolation. In a similar interpretative vein, Coppola and Masullo (2021) recognise the virtual environment, specifically Facebook, as a means for hikikomori of maintaining a form of social identity, sharing with the other members of the community their feelings, thoughts and beliefs. Results reported in Cavalieri and Corrizzato's linguistic study (2025) also explain that Reddit is seen as a virtual space in which hikikomori have the possibility to interact and describe their feelings without the pressure given by offline society.

On the other hand, other studies question the interactional validity of these exchanges. While some online communication may involve meaningful contact with other individuals, much of it occurs in mediated or semi-anonymous environments, e.g., through interactions with avatars or participation in online gaming, where interactants may never engage in direct interpersonal contact with the risk to reinforce isolation (Esposito and Addeo 2025; Fong and Yip 2023; Gavin and Brosnan 2022).

Considering both the advantages and the limitations of digital interaction, the present contribution adopts an intermediate stance: although online fora inevitably lack many of the relational characteristics of face-to-face communication, they can provide a meaningful arena in which users share and develop their ideas and engage with like-minded peers. In this regard, a linguistic analysis of the written discourse can provide a valuable tool for going beyond the surface level of the text, uncovering meanings that can shed light on the interlocutors' identities and interactional styles.

Building on this brief introduction, the present study aims to fill a gap in the literature by providing a linguistic perspective on hikikomori, an angle still understudied in favour of psychiatric and sociological approaches since Saito's pioneering studies (1998). Specifically, examining how hikikomori describe themselves, their environment, and their emotions in virtual spaces can provide useful insights into the subjective dimensions of the phenomenon.

To this objective, the next paragraph will describe the role of emoticons and emojis in contributing to the pragmatic force of utterances, whereas in section 4 the methodological framework of our study will be presented. In the next paragraphs, the quantitative and qualitative results will be included and contextualised.

2. Emoticons and emojis as pragmatic devices

Emoticons, a blend of “emotion” and “icon”, are pictorial representations of emotions using keyboard characters, such as the classic smiley face – :-). Their primary function was to allow users to indicate their emotions more accurately (e.g. Dresner and Herring 2010), particularly in light of missing contextual cues naturally present in face-to-face interactions, that is, verbal and non-verbal features such as suprasegmentals, facial expressions and gestures. Within online environments – blogs, discussion fora, social media platforms, and (mobile) instant messaging applications among others – most communication occurs in written form. The use of emoticons contributed to greater clarity and nuance in online exchanges. Indeed, Huang *et al.* 2008)’s study revealed that users perceived emoticons as a useful tool to convey additional information, making text-based interactions richer and more expressive.

As technology progressed and new affordances were introduced in internet-mediated communication environments, emoticons evolved from combinations of keyboard characters to more sophisticated small pictorial icons named emojis – from the Japanese *e* (絵), “picture,” and *moji* (文字), “character.” These comprise a much wider range of representations than emoticons, including facial expressions, people, abstract concepts, animals, plants, activities, travel (Li and Yang 2018; Rodrigues *et al.* 2017). The repertoire of emojis has gradually expanded over time: the full emoji list now contains 3953 emojis, including gender and skin variations.²

Multiple studies have explored the semantic meaning of individual emojis in isolation (e.g. Jaeger and Ares 2017); while these perception studies provide valuable insight, it is equally significant to examine how emojis are employed within authentic interactions. Their interpretation is indeed highly context-dependent: the same picture may be understood or used differently within different communities and cultures (Bettelli and Panzeri 2023; Gibson *et al.* 2018; Widlana *et al.* 2024), where the use of emojis was determined to be – to different degrees – influenced by the local culture in either meaning or context of use.

Yus (2014) stresses this point by theorizing a continuum that extends from purely arbitrary to explicitly iconic interpretations, driven by cultural norms for decoding visual signs. This continuum illustrates that understanding the meaning of an emoji often requires some form of textual “anchorage” or contextual grounding within the message. In other words, the

² <https://unicode.org/emoji/charts/full-emoji-list.htm> (19.09.2025).

meaning of an emoji is seldom fixed and instead relies heavily on the surrounding text and the social or cultural context in which it is used.

Beyond their well-established role in adding accuracy to expression of emotion, emojis also contribute to the construction of both social identity and meaning, serving a variety of pragmatic functions in communication, either alone or alongside text. Indeed, studying internet-mediated interactions without considering the visual information provided by emojis “would lead to the loss of important information that conveys people’s feelings or clarifies the meaning of the text” (Kusal *et al.* 2025, p. 13). The importance of emojis is also highlighted by research that suggests that their use may influence how readers perceive the emotionality and clarity of a message as well as sender warmth (*et al.* 2022).

Yus’s concept of Cyberpragmatics (2011) is often employed as a theoretical background to analyse how meaning is constructed in internet-mediated interactions where paralinguistic and extralinguistic cues are limited or entirely absent. Cyberpragmatics, as the name suggests, focuses on the distinct characteristics of digital exchanges, considering how text deformation,³ emoticons and emojis, and other audio-visual aspects of internet-mediated communication function pragmatically in these interactions.

Research on emoji is however not limited to linguistics and communication studies: other disciplines, such as psychology, have also looked at communication in a variety of digital environments, including blogs, social media platforms like Facebook and Instagram, and instant messaging clients and apps. These studies have revealed the multifaceted roles of emoticons and emojis in internet-mediated interactions.

Emojis have been shown to be employed as politeness strategies, where they are used as face-saving devices (e.g. Togans *et al.* 2021), or to help mitigate face-threatening acts (e.g. Kavanagh 2016). Related to their role in politeness, emojis also play an important role in rapport-building, contributing to managing face, turn-taking, and as markers of informality (e.g. Sampietro 2019). In (mobile) instant-messaging, in addition to turn-taking devices, emojis may also act as a form of backchanneling (e.g. Li and Yang 2018). Expression of humour and irony has also been attested as a function of emoji (e.g. Bettelli and Panzeri 2023).

Multiple taxonomies for the pragmatic functions of emojis have been developed, from simpler, three-way classifications (e.g. Dresner and Herring 2012; Luor *et al.* 2010) to more articulated categorisations, such as Kavanaugh’s (2010), Yus’s (2014), and Li and Yang’s (2018), which builds

³ Text deformation includes different strategies such as use of capitalization, repetition of letters, or alternative uses of punctuation marks for the purpose of verbally expressing emphasis, attitude and feelings in textual internet-mediated interactions.

on Yus's taxonomy. The revision simplified function names for increased comprehensibility, combined functions with foggy borders, and added a turn-taking and backchanneling device to the list of functions. The authors applied their taxonomy to a corpus of socio-emotional communication on the Chinese instant-messaging service *WeChat*. Socio-emotional communication settings appear to show more frequent use of emojis than task-oriented exchanges; we may posit that an online community of hikikomori is a social-emotional setting, and while Reddit allows for the sharing of pictures and gifs in comments, they are not as widely used as emojis. For the purpose of this study, we will adopt a revised version of their categorisation, as will be further detailed in the next section.

3. The corpus

For the present study, the analysis was carried out on a corpus of Reddit posts, hereafter referred to as the Hikikomori Reddit Corpus. Reddit, launched in 2005 by Steve Huffman and Alexis Ohanian, is a social news and discussion website that has become increasingly popular over the last two decades. On this sharing platform, user-generated content is posted in interest-based communities called subreddits, signalled by the prefix *r/*. Subreddits may cover a wide variety of topics, ranging from minority groups and fan communities to online phenomena such as memes and current events (Yadav *et al.* 2021, p. 1286).

According to the most recent statistics (Q2 2025), Reddit hosts more than 100,000 active communities (subreddits) and has surpassed 1 billion cumulative posts since its launch. The platform counts approximately 110 million daily active users (DAU) and over 416 million weekly active users (WAU), marking a significant increase compared to previous years.

Reddit functions similarly to an online forum: registered users publish posts (also called *threads*) in subreddits, where they share items (text, pictures, videos, links, etc.) for other community members to comment on and discuss. Anonymity is a distinctive feature of Reddit, as users are not required to disclose personal identity or location. Each subreddit has its own guidelines, maintained by moderators to ensure respectful participation. Another defining characteristic is Reddit's voting system, whereby users can upvote or downvote posts and comments. Content with higher numbers of upvotes gains more visibility within the subreddit and the site as a whole (Michellebrous 2017).

The *Hikikomori Reddit Corpus* was compiled by collecting all posts published on the subreddit *r/hikikomori* between January 2020 and July 2023,

amounting to approximately 2 million words.⁴ The starting point in 2020 was chosen to capture discussions in the aftermath of the Covid-19 pandemic, which has been reported to exacerbate conditions of social withdrawal. Within this subreddit, individuals living with the hikikomori condition can find a virtual “place” to discuss their situation of psychological and social isolation.

The data was retrieved using a Python script from the PushShift archive, a service that gathers and stores Reddit submissions and comments. Two batches of data were collected: one consisting of submissions, the other of comments. Both batches underwent anonymisation: all usernames were replaced by unique identifiers composed of the prefix “id” followed by six random digits (e.g., *id679841*). The resulting datasets were stored as CSV files, with non-relevant columns dropped in order to retain only the anonymised author, the text (*selftext*), posting time (*created_utc*), and other attributes varying by post type. Posts with no textual content (*selftext*), as well as those marked as [deleted] or [removed], were excluded. The temporal metadata originally stored as Unix epoch time was reformatted to a standard *YYYY-MM-DD HH:MM:SS* UTC format. Each batch was also labelled with a *type* column (submission vs. comment), to facilitate the creation of subcorpora. The two datasets were then merged into a single file and converted into XML, a format suitable for corpus linguistic tools, where metadata annotation (e.g., author, date, type) is kept distinct from the textual fields (title and body).

As a further step, the XML corpus underwent emoji tagging in order to enable pragmatic analysis. Each emoji was automatically detected through Unicode-based regular expressions, then normalised into canonical forms (e.g., coloured hearts → HEART, stars → STAR, skin tone modifiers removed). Normalised tokens were wrapped in dedicated XML tags (e.g., `<emoji type="HEART"/>`) so that they could be queried and analysed in corpus platforms such as SketchEngine. This tagging process made it possible to retrieve emoji systematically, to treat them as searchable units parallel to words, and to connect them to their pragmatic functions according to the coding scheme adopted in this study (Li, Yang 2018; Yus 2011, 2021).

4. Methodology

Drawing on Corpus Linguistics and cyberpragmatics, the study adopted a mixed-method approach and data was investigated both from a quantitative

⁴ We would like to thank Hunter Paul Youngquist, who cooperated in compiling the Hikikomori Reddit Corpus.

and a qualitative perspective. In the first phase of the investigation, emojis were automatically parsed from the corpus using ElementTree, which is a Python XML parsing library. To extract the emojis carefully, two types of emotional expressions were considered: contextualised emojis in XML tags and Unicode emoji codes. The corpus was read in a systematic way to access all text contents, using rather flexible parsing methods in order to retain data content despite possible structural inconsistencies within the file format. All the extracted emojis were collected and then analysed with the Counter class (Python 3) to calculate frequency distribution. At the end of this extraction, we focused on the ten most frequent emojis found in our corpus and we looked at their context of use through the concordance tool of *SketchEngine* (Kilgarriff *et al.* 2014) to pinpoint the predominant pragmatic functions for the subsequent qualitative analysis.⁵

For the qualitative pragmatic analysis of emojis, we adopted a functional categorisation based on a framework that integrates Li and Yang's (2018) corpus-based taxonomy with Yus's (2011, 2021) cyberpragmatic categories. The following table provides an overview of these categories, outlining the main functions they perform in online interaction.

Li and Yang (2018)	Integrated description
Attitude / Emotion signals	Express emotions and attitudes; build empathy and solidarity.
Intensity enhancers	Reinforce emotional intensity; boost pragmatic salience of the utterance.
Humour / Irony	Orient interpretation toward irony, humour or sarcasm.
Illocutionary force modifiers	Modify or soften the illocutionary force; disambiguate speaker intention.
Turn-taking / Backchannel signals	Maintain/open the channel; signal acknowledgement or minimal agreement.
Identity markers (Yus 2011, 2021)	Emoji as markers of belonging, self-positioning or stylistic identity.

Table 1
Framework for the pragmatic functions of emojis.

In the sections that follow, the findings are reported beginning with a quantitative overview and moving on to a detailed qualitative analysis.

⁵ We would like to thank Javad Haditaghi, who helped us with the Python code for extracting and calculating the emojis in the corpus.

5. Results and discussion

5.1. Quantitative analysis: focus on the top ten emojis in the r/Hikikomori corpus

In order to provide an initial quantitative overview of emoji presence in the Hikikomori Reddit corpus, Table 2 reports the ten most frequent items together with the percentages of the pragmatic functions they are associated with:

Emoji	Frequency (% of all emojis)	Main pragmatic functions (with % of total uses)
😄	4.68	Express emotions & attitudes (49.5%); Modify/soften the illocutionary force (14.0%); Irony/humour (12.9%); Boost salience (12.9%)
😂	3.66	Irony/humour (76.8%); Build empathy & solidarity (15.9%); No clear meaning (4.5%)
🤔	3.27	Modify/soften illocutionary force (48.4%); Maintain/open the channel (19.4%); Boost salience (14.5%); Reinforce emotional intensity (9.7%)
👍	2.88	Boost salience (42.6%); Reinforce emotional intensity (24.1%); Express emotions & attitudes (17.0%); Acknowledgement/minimal agreement (11.1%)
😊	2.19	Express emotions & attitudes (46.7%); Reinforce emotional intensity (35.6%); Build empathy & solidarity (17.8%)
😓	2.19	Express emotions & attitudes (25.6%); Build empathy & solidarity (17.9%); Acknowledgement/minimal agreement (25.6%); Irony/humour (12.8%)
😭	2.15	Reinforce emotional intensity (94.4%); Irony/humour (5.6%)
😬	2.15	Irony/humour (90.6%); Build empathy & solidarity (6.3%); No clear meaning (3.1%)
🙏	2.05	Express emotions & attitudes (65.7%); Build empathy & solidarity (34.3%)
❤️	1.90	Express emotions & attitudes (40.0%); Reinforce emotional intensity (20.0%); Build empathy & solidarity (14.3%); No clear meaning (2.9%)

Table 2

Top ten emojis in the *Hikikomori* corpus: frequency and main pragmatic functions.

As is possible to observe from the table, each emoji may be associated to different pragmatic functions and sometimes one emoji can perform two functions at the same time. For the purpose of our study, however, we have decided to focus only on the primary pragmatic function enacted by each emoji, rather than on secondary or concurrent functions, that is, the function most prominently and transparently conveyed by the emoji in context. The categorisation was based on full agreement among the three researchers.

The most recurrent emoji is the Grinning Face with Sweat 😄 (4.68%), which displays a heterogeneous functional profile. As the distribution in our corpus shows, it is employed primarily to *express emotions and attitudes* (49.5%) but also contributes to *mitigating illocutionary force* (14.0%), *irony and humour* (12.9%), and *boosting salience* (12.9%).

The second most frequent emoji, 😄, Face with Tears of Joy, (3.66%), has a much more “specialised” use and it is almost exclusively used to *orient interpretation toward irony and humour* (76.8%), with additional contributions to *solidarity* (15.9%). This profile confirms its centrality in reinforcing collective sharing within humorous alignment, in line with Yus’s (2011, 2021) account of affective intensification. Similarly, 🤣, Rolling on the Floor Laughing, (2.15%) is almost exclusively humorous (90.6%), marking amusement as also a means of solidarity in the Hikikomori group.

By contrast, the Thinking Face 🤔 (3.27%) has a different orientation. It functions primarily to *modify or soften illocutionary force* (48.4%), and to *maintain the channel* (19.4%), while also slightly contributing to *salience boosting* (14.5%). This emoji is less about affect and more about guiding the interpretation of the utterance. Thumbs Up 👍 (2.88%), on the other hand, is mainly used to *boost salience* (42.6%), *reinforce intensity* (24.1%), thus it functions as a marker of digital backchannel sustaining conversational continuity.

The relational orientation is particularly evident in the Smiling Face with Smiling Eyes 😊 (2.19%) and the folded hands 🙏 (2.05%). The former is used to *express positive emotions* (46.7%), but also to *reinforce intensity* (35.6%) and *build solidarity* (17.8%). The latter is more solidarity-driven as it appears mainly in *expressive gratitude* (65.7%) and *solidarity-building* contexts (34.3%). Similarly, the Red Heart ❤️ (1.9%) is used across several categories, though it is especially salient in *express emotions* (40.0%) and *reinforce intensity* (20.0%), consolidating affective ties in supportive exchanges.

Interestingly, the Beaming Face with Smiling Eyes 😁 (2.19%) fulfills different functions: it contributes to *expressing emotions* (25.6%), it is equally used to *signal acknowledgement* (25.6%) and, to a lesser extent, to *build solidarity* (17.9%) and *mark humour* (12.8%). Finally, the Loudly Crying Face 🤔 (2.15%) is highly “specialised”, with 94.4% of its uses reinforcing *emotional intensity*, thus dramatising affective stances and personal narratives, in line with the expressive amplification highlighted in previous literature.

Macrocategory	Pragmatic Function (with representative emojis)	Percentage (%)
Attitude / Emotion signals	Express emotions & attitudes 😄 😊 😁 🙏 ❤️	24.5
	Build empathy & solidarity 🤔 😊 😁 🙏 ❤️	12.0
Intensity enhancers	Reinforce emotional intensity 🤔 👍 😊 🤔 ❤️	15.9
	Boost pragmatic salience 😄 🤔 👍	10.4
Humour / Irony	Irony / humour / sarcasm 🤔 🤣 😊 🤔 🤔	20.3
Illocutionary force modifiers	Modify / soften the illocutionary force 🤔 🤔	2.6

Turn-taking / Backchannel signals	Maintain / open the channel 🙄 😊	9.4
	Acknowledgement / minimal agreement 👍 😊	3.0
No clear meaning		2.0

Table 3
Distribution of pragmatic functions and associated emojis.

Overall, as shown by Table 3, which represents the general distribution of the pragmatic functions throughout the dataset, emojis in our corpus are predominantly used for expressive and affective purposes, i.e. the categories *express emotions and attitudes*, *irony and humour*, and *reinforce emotional intensity*, which account for over 60% of all occurrences. This pattern may reflect a tendency of the Hikikomori community to rely on emojis as affective/relational devices rather than propositional markers in line with what is assumed by Li and Yang (2018) and Yus (2021) on the function of emojis as markers of interpersonal cohesion. Moreover, humour- and irony-related functions (20.3%) constitute the second dominant area, aligning with Yus's (2021) notion of cyberpragmatic play, i.e. emojis help speakers frame potentially delicate experiences in self-ironic ways, softening vulnerability through shared amusement. Functions such as *boost pragmatic salience* (10.4%) and *maintain/open the channel* (9.4%) further point to the interactional scaffolding role of emojis, ensuring continuity and mutual involvement.

In contrast, categories such as *modify/soften illocutionary force* (2.6%) and *acknowledgement/minimal agreement* (3.0%) have a lower presence, signalling that politeness or mitigation play a secondary role compared to affective engagement.

The next section turns to the qualitative analysis, illustrating each pragmatic macrocategory with representative examples from the corpus.

5.2. Qualitative analysis: focus on the pragmatic functions of emojis

5.2.1. Attitude / Emotion signals

In the Hikikomori Reddit Corpus, the macro-category *attitude/emotion signal* encompasses emoji that foreground affective stance and interpersonal orientation. Within this broader category, two subtypes emerge as particularly significant in our data, i.e. *expressing emotions and attitudes* and *building empathy and solidarity*. The first subtype includes emojis such as 😊, 😊, ❤️, 🙏, and 😊, which visually encode affective states such as gratitude,

embarrassment and affection. The second features the same emojis, but with the function of fostering solidarity and empathetic connection among users.

A closer analysis of the *expressing emotions and attitudes* subtype reveals that these emojis frequently function as affective expressions, they act as “metarepresentational level of information beyond the explicit content of the utterance” (Yus 2014, p. 7). In example (1) below, the user employs the folded-hands emoji to express gratitude by visually enacting thankfulness:

- (1) I appreciate the response 🙏 I thought i was taking small steps but idk I think the small store you were talking about sounds like a good idea and seems

Similarly, in example (2), we find another emoji with the same function; the smiling face emoji underscores friendliness and warmth, adding emotion to the gratitude message:

- (2) Thank you 😊 Boa sorte amigo

The user’s intent here is not merely to acknowledge but to establish warmth and goodwill, particularly in intercultural or multilingual exchanges.

Within *express emotions and attitudes*, 😓 recurrently indexes embarrassment and light self-deprecation. Example 3 pairs apology with a smile-with-sweat cue, visually indicating a embarrassment for what is stated before the emoji:

- (3) I’m sorry if that’s a bad thing 😓

This example highlights the dual function of emojis in both conveying and modulating emotional disclosure, enabling users to navigate a delicate balance between authentic self-expression and interpersonal self-preservation. The 😓 emoji was often used to express a self-deprecating attitude, especially in relation to the social challenges faced by hikikomori. This suggests awareness of their status and a certain degree of embarrassment on the part of certain users at their inability to function as full-fledged members of society.

These uses align with Li and Yang’s (2018) expressive category, where emojis function as intensifiers that reinforce verbal stance while shaping interpersonal dynamics. Moving to the *building empathy and solidarity* subtype, we observe that emojis contribute not only to affective expression but to relational alignment, fostering shared emotional ground. In example (4), the folded-hands emoji extends beyond gratitude to perform a gesture of solidarity in response to personal disclosure:

- (4) Hope things get better 🙏 I'm also in my 30s, also Asian, also a college drop-out

The emoji bridges the gap between the speakers signalling shared experience as a foundation for support. This logic of emotional reciprocity is further illustrated in example (5), where the smiling emoji subtly reframes an offer of friendship, making it appear caring and non-intrusive:

- (5) Same. we can be friends, I'm lonely a lot too 😊

The emoji here downgrades the seriousness of the statement, facilitating a safe environment for relational outreach. Finally, in example (6), the grinning face emoji infuses the message with a tone of encouragement:

- (6) Take care 😊 I feel you man

Together, these examples illustrate what Maíz-Arévalo and García-Gómez (2013: p. 741) describe as “relational work” in digital discourse – emojis operate as multimodal resources that mitigate social distance and create bonds of mutual understanding.

To sum up, the *attitude/emotion signal* category does not simply reflect internal emotional states; rather, it enables users to perform socially situated affect, negotiating both stance and affiliation. As Yus (2021) argues, the pragmatic potential of emojis lies not in their semantic content but in their capacity to manage interpersonal meaning. Within the *Hikikomori* subreddit in which expressions of isolation and emotional fragility are common, this dual function of expressing and co-constructing emotion becomes especially salient, supporting solidarity and empathy within its members.

5.2.2. Intensity enhancers

The emojis under investigation were also commonly used as intensity enhancers, reinforcing either the emotional intensity or the pragmatic salience of the proposition. In these contexts, the user's proposition is expressed verbally, and the emoji is added to further reinforce or intensify the propositional or affective attitude of the speech act (Yus 2014, p. 520, cf. also Li, Yang 2018; Sampietro 2019). Out of the 10 emoji, 8 were used to fulfill this function in our data, for a total of 132 instances.

In example (7) below, the user employs the Grinning Face with Sweat emoji in combination with the angel emoji.

- (7) Sorry for my inactivity again, but plz know if you guys need anything don't be afraid to reach out. 😇😓

The 🙄 emoji here seems to upgrade the pragmatic salience of the apology for not being active on the subreddit, also expressing awkwardness at their lack of presence. The Smiling Face with Halo emoji, not under investigation in this study, appears to play a dual function here, acting as an intensifier of their offer to help, and possibly as an irony marker, in contrast with his undesirable behavior, further highlighting the apology in the proposition.

In the following example, the Thinking Face emoji also boosts the pragmatic salience of the proposition, reinforcing both the contemplation of the statement they are responding to and the skepticism of the ensuing question.

- (8) Wait 🤔 If your dad kicked you out and you spend your money on traveling, then how do you support yourself now?

The Thinking Face may also express doubt and skepticism, and its use here introduces the verbally-coded proposition that threatens the other user's face by questioning the truth of their claims.

In example (9) below, the Smiling Face with Smiling Eyes emoji (😊) is used to both enhance the thanking speech act and the affective attitude of the author, representing gratitude and warmth.

- (9) NF for example I am so happy to hear that you managed to find a job! What you are doing here trying to help us hiki is awesome ✨😊

The presence of the Sparkles emoji alongside 😊 further amplifies the affective attitude of the message. This emoji not only contributes to expressing gratitude but also expresses excitement and a sense of celebration regarding the addressee's newfound job.

These examples show that often emojis do not fulfil a single, clear-cut function in messages, but they may provide additional visual cues to verbal propositions, adding detailed nuance to online interactions.

5.2.3. *Humour / Irony*

In the framework of cyberpragmatics, laughter-related emoji such as 😂 and 🤣 play a crucial role as paralinguistic cues which both frame and reinforce humorous intent in digital interaction (Dynel 2016; Yus 2021). Online discourse, indeed, lacks the prosodic and kinesthetic signals which are traditionally part of face-to-face communicative exchanges. Hence, laughter-related emojis, 😂 (occurring 53 times in this function) and 🤣 (29 times), serve to indicate that the message should be read with a humorous undertone. In this way, they guide community members to interpret the utterance non-literally, as the speaker intentionally flouts conversational maxims to signal a

non-bona-fide mode of communication (Raskin 1985).

Our corpus reflects this tendency, as the two emoji are included when the interlocutors want to reinforce the cooperative alignment among the members of the community, building or maintaining rapport in digital forum. In the following example, the author of the message reports on their renewed inclination to drive and provides a humorous account of the reactions of their neighbours upon seeing them behind the wheel.

- (10) The young kids in the neighbourhood were watching me in the car driving and talking about me, 🤔. Some of them probably wasn't even born when I first started driving, so this is their first time seeing me do adult things.

Ideally, 😂 and 😄 also illustrate the gradient of intensity in affective stance: 😂 often functions as a marker of recognition or affiliative laughter, whereas 😄 indexes heightened amusement or exaggerated hilarity (Provine 2000; Danesi 2017). In this sense, our data do not align with these theories, as the two emoji coexist and do not appear to be selected according to the level of amusement.

Consistent with prior studies, our research demonstrates, however, that the two emojis are also employed when interlocutors intend to convey irony, marking it through their use. In the context of the hikikomori community, irony thus emerges as a self-reflexive device through which individuals narrate their experiences and behaviours in a mocking manner, positioning themselves at once inside and at a distance from their own stories. In the following posts, irony arises from the incongruity between normative expectations and the two hikikomori's lived realities. In the first, the suggestion to travel abroad contrasts humorously with the individual's difficulty in leaving their room, while in the second, temporary online interactions are ironically framed as equivalent to friendships.

- (11) A person recommended I travel to Japan, and I laughed in abysmal despair. Like I'm afraid to attend college, let alone leave my own room. I can dream, tho! 😂
- (12) Welcome to the club 🤔 People actually make friends here? yep. sometimes it lasts 6hrs, usually 12hrs and if i'm lucky i get a whole day 🤔 having relationships with fictional characters is pretty normal.

Within the corpus, additional emojis are employed to mark humorous utterances, although with considerably lower frequency; notably, 🤔 occurs only 12 times in this function. In the following post, for instance, the inclusion of 🤔 functions as a pragmatic cue, mitigating the potentially face-threatening force of the expression and framing it instead as playful self-irony. In this way, the utterance both acknowledges a stigmatized desire

within the hikikomori community and reinterprets it through humorous stance-taking, guiding recipients toward a non-literal meaning.

(13) Yeah, no shit I think all of us wish to be normal, lol 🤔

5.2.4. Illocutionary force modifiers

Emoticon and emoji use has been linked to enhancing or softening the illocutionary force of the speech act they are attached to, expressing user intentions in a more accurate manner and clarifying the meaning of their words (Yus 2014). Prior research has shown (e.g., Sampietro 2018; Kavanagh 2016; Yus 2014; Dresner, Herring 2010) that emoticons and emojis are employed to downgrade or mitigate the impact of directive speech acts that may otherwise be perceived as face-threatening by readers. This strategy was however not frequently associated to the emojis under investigation in our corpus: only 13 instances were identified, all of which involved the Grinning Face with Sweat Emoji (😓). This emoji is often defined as expressing embarrassment or awkwardness, relief after a stressful situation, or again, management of a stressful situation, as may be seen in the examples below.⁶ It may also be used to express self-criticism or self-deprecation, as was seen in section (5.2.1).⁷

(14) Lol Doom scrolling social media. That sounds absolutely horrid 🤔
Perhaps play Minecraft?

In this context, the author of the comment is reacting to another user's complaints about their online gaming experience, which they compare to 'doom scrolling' on social media. The use of the emoji here mitigates the strength of the adjective 'horrid', signalling the light, almost humorous tone of the comment, while at the same time expressing awkwardness at the other user's negative experience.

In the following example, the emoji is used to soften the face-threat of a statement rather than a directive, and to mark the proposition as a stressful situation that many people, not just hikikomori, have to deal with:

(15) Maybe not hikis specifically, but I guarantee there's a bunch of lonelyass introverted people sitting in their rooms alone feeling the same way 🤔 So many people are alone and want love & friendship

⁶ <https://emojiterria.com/smiling-face-with-open-mouth-cold-sweat/>;
<https://englishonline.britishcouncil.org/blog/articles/beyond-a-smiling-face-your-guide-to-uncommon-emoji-meanings/> (27.09.2025).

⁷ <https://emoji-graph.org/grinning-face-with-sweat/> (27.09.2025).

Here the author replies to another user complaining that they may not meet any other hikis in their town to bond with. He corrects their assumption pointing out that loneliness is a common condition and they are likely to find someone looking for human contact. The emoji here, as stated above, appears to play a dual purpose: on the one hand, it mitigates the face-threat of the disagreement expressed in the message; on the other, it highlights loneliness as stressful condition that is shared by many in contemporary society.

While other, less frequent emojis may have been used in the subreddit to modify or soften the illocutionary force of speech acts, it appears that this function is not as common here. Verbal downgraders and mitigators may be preferred or, as an informal support community, directive speech acts such as exhortations may not be mitigated. Further research would help clarify the role of directives in such communities.

5.2.5. *Turn-taking / Backchannel signals*

Within our corpus, the following emojis 🤔 (14 occurrences), 🤔 (12 occurrences) and 👍 (6 occurrences) also function as backchanneling devices, serving to maintain or initiate the communicative channel and to signal acknowledgement or minimal alignment, as the following examples show:

(16) Thats like our holy national dish , u can buy those dishes almonst everywhere. That guy sucks as a t guide, if he failed to even mention about momo. U taste nepal when u eat it 🤔 and ull come back for more Im sure ull go to bhutan one day, plz try momo there.

(17) I think you've got it backwards: if you were literal trash, someone else would have to throw you away. So, rejoice in your capability.
 ✨🤔👍✨

In the first post the emoji 🤔 does not introduce new propositional content but reinforces the speaker's enthusiastic stance, thereby sustaining the positive tone of the exchange, confirming signal acknowledgement. Similarly, the 👍 emoji in the second post, together with the other ones, functions primarily as a phatic marker, signalling acknowledgement too. In both cases, the emojis operate as pragmatic resources that guide interpretation and contribute to enhancing cooperation among the interactants.

6. Concluding remarks

This study has explored the pragmatic functions of emojis in the r/hikikomori subreddit, a digital space where individuals experiencing extreme social withdrawal communicate and construct shared meaning. By integrating corpus linguistics with a cyberpragmatics framework, the research has illustrated how emojis serve not merely decorative purposes but perform pragmatic functions related to the expression and modulation of emotion, identity, and interpersonal stance and community dynamics. These results are in line with Li and Yang's 2018 study, which highlighted how emojis are primarily socio-emotional tools.

The quantitative analysis revealed that a core group of ten emojis dominates interaction within the community, with each emoji demonstrating a range of pragmatic functions. Notably, emojis such as 😓, 😂, and 🤔 perform multifunctional roles, i.e. conveying embarrassment, signalling humour or softening illocutionary force. This could mirror the nuanced affective landscape of hikikomori discourse due to the varied community of the sub-reddit.

From a qualitative perspective, the analysis of emoji use also reveals that hikikomori's inclusion of emoji serves as an important resource, fostering cooperation and developing a sense of collective belonging within the digital community. Although the members do not know each other personally, emojis help them to share their experiences and emotional states and to express affective alignment, reinforcing group cohesion in computer-mediated communication. Predictably, the turn-taking and backchannelling functions were rare, as Reddit is an asynchronous platform, where many people interact by starting different threads in response to user posts. It is a different context from instant messaging where minimal agreements and backchannel devices would be more common (Li, Yang 2018).

While the present analysis focused on the most frequently used emojis, this perspective inevitably limited the exploration of more creative uses, including their (potential) role as identity markers and the emergence of alternative, community-based meanings. This methodological choice was due to the large size of the corpus, which made a more fine-grained analysis difficult to manage. Future research could therefore benefit from adopting a more qualitative approach on a smaller sample of the corpus, allowing experts to investigate how emojis function as resources for identity construction in such a community.

Moreover, future research could explore text-based emoticons, which were excluded from the present analysis due to challenges in parsing complex character combinations, as well as examine hikikomori individuals' use of emojis across other online platforms. Such analysis could offer comparative insights into whether and how different online fora can shape patterns of

emoji use.

In conclusion, the findings of this study suggest that emojis should not be understood merely as tools for expressing emotion but rather as elements of a broader communicative repertoire. Within this context, emojis appear to function as a form of language, contributing to the construction of politeness strategies, interpersonal cohesion, and group identity. By examining emoji use, this analysis highlights how this visual code actively participates in shared norms and meaning-making practices, offering insight into the social and pragmatic dimensions of digital communication.

Bionotes: Silvia Cavalieri is Associate Professor of English Linguistics and Translation at the University of Modena and Reggio Emilia. She has worked and published on various aspects of discourse analysis and ESP with a focus on corpus linguistics and digital humanities. Her most recent works have centred on different specialized discourses in the digital environment, especially considering social media and user-generated contents.

Sara Corrizzato is Associate Professor in English Language, Translation and Linguistics at the Department of Foreign Languages and Literatures of the University of Verona. Her current areas of research focus on corpus linguistics and Digital Humanities, ESP and audiovisual language. She is currently working on two projects: the analysis of the language used in spoken interactions by diplomats and international operators, and the study of the linguistic strategies adopted by professionals to disseminate scientific concepts to non-experts.

Valeria Franceschi is Associate Professor of English Language, Translation and Linguistics at the Department of Foreign Languages and Literatures of the University of Verona, where she teaches at undergraduate and graduate level. Her research interests include English as a Lingua Franca, with a focus on plurilingualism and BELF, tourism discourse, and more recently, science popularisation discourse, which she analyses primarily with a corpus-assisted approach.

Authors' addresses: silvia.cavalieri@unimore.it, sara.corrizzato@univr.it,
valeria.franceschi@univr.it.

References

- Bettelli G. and Panzeri F. 2023, *"Irony is easy to understand": The Role of Emoji in Irony Detection*, in "Intercultural Pragmatics" 20 [5], pp. 467-493.
- Cavaliere S. and Corrizzato S. 2025, *"Now, I have a job and my life started improving": A Linguistic Analysis of Former Hikikomori's Perspective on the World of Work*, in Tommaso L. and Zummo M.L. (ed.), *Exploring New Occupational Discourses and Identities across Genres: Crisis and Well-being*, Cambridge Scholars Publishing, Newcastle Upon Tyne, pp. 168-185.
- Coppola M. and Masullo G. 2021, *The COVID-19 Pandemic through the Eyes of Italian Young Hikikomori*, in "Culture e Studi del Sociale" 6, pp. 219-226.
- Dresner E. and Herring S.C. 2010, *Functions of the Nonverbal in CMC: Emoticons and Illocutionary Force*, in "Communication Theory" 20 [3], pp. 249-268.
- Esposito V. and Addeo F. 2025, *Hikikomori 2.0: A Mixed Method Investigation to Reconceptualization the Social Withdrawal in the Digital Age*, in "Societies" 15 [9], no. 260.
- Fong T.C. and Yip P.S. 2023, *Prevalence of Hikikomori and Associations with Suicidal Ideation, Suicide Stigma, and Help-seeking among 2022 Young Adults in Hong Kong*, in "International Journal of Social Psychiatry" 69, pp. 1768-1780.
- Gavin J. and Brosnan M. 2022, *The Relationship between Hikikomori Risk and Internet Use during COVID-19 Restrictions*, in "Cyberpsychology, Behavior, and Social Networking" 25, pp. 189-193.
- Gibson W., Huang P. and Qianyun Y. 2018, *Emoji and Communicative Action: The Semiotics, Sequence and Gestural Sections of "Face covering hand"*, in "Discourse, Context & Media" 26, pp. 91-99.
- Hand C.J., Burd K., Oliver A. and Robus C.M. 2022, *Interactions between Text Content and Emoji Types Determine Perceptions of both Messages and Senders*, in "Computers in Human Behavior Reports" 8, no. 100242.
- Huang A.H., Yen D.C. and Zhang X. 2008, *Exploring the Potential Effects of Emoticons*, in "Information & Management" 45 [7], pp. 466-473.
- Jaeger S.R. and Ares G.G. 2017, *Dominant Meanings of Facial Emoji: Insights from Chinese Consumers and Comparison with Meanings from Internet Resources*, in "Food Quality and Preference" 62, pp. 275-283.
- Kato T.A., Kanba S. and Teo A.R. 2019, *Hikikomori: Multidimensional Understanding, Assessment, and Future International Perspectives*, in "Psychiatry and Clinical Neurosciences" 73 [8], pp. 427-440.
- Kavanagh B. 2010, *A Cross-cultural Analysis of Japanese and English Non-verbal Online Communication: The Use of Emoticons in Weblogs*, in "Intercultural Communication Studies" 3, pp. 65-80.
- Kavanagh B. 2016, *Emoticons as a Medium for Channeling Politeness within American and Japanese Online Blogging Communities*, in "Language & Communication" 48, pp. 53-65.
- Krieg A. and Dickie J.R. 2013, *Attachment and Hikikomori: A Psychosocial Developmental Model*, in "International Journal of Social Psychiatry" 59 [1], pp. 61-72.
- Li L. and Yang Y. 2018, *Pragmatic Functions of Emoji in Internet-based Communication: A Corpus-based Study*, in "Asian-Pacific Journal of Second and Foreign Language Education" 3, no. 16.
- Li T.M.H. and Wong P.W. 2015, *Editorial Perspective: Pathological Social Withdrawal during Adolescence: A Culture-specific or a Global Phenomenon?*,

- in "The Journal of Child Psychology and Psychiatry" 56 [10], pp. 1039-1041.
- Luor T., Wu L.-L., Lu H.-P. and Tao Y.-H. 2010, *The Effect of Emoticons in Simplex and Complex Task-oriented Communication: An Empirical Study of Instant Messaging*, in "Computers in Human Behavior" 26 [5], pp. 889-895.
- Maíz-Arévalo C. and García-Gómez A. 2013, "You look terrific!" *Social Evaluation and Relationships in Online Compliments*, in "Discourse Studies" 15, pp. 735-760.
- Michellebrous. 2017, *Reddit: Managing the "Front Page of the Internet"*, in *Digital Innovation and Transformation*. <https://digital.hbs.edu/platform-digit/submission/reddit-managing-the-front-page-of-the-internet/> (01.10.2025).
- Neoh M.J.Y., Carollo A., Lim M. and Esposito G. 2023, *Hikikomori: A Scientometric Review of 20 Years of Research*, in "International Journal of Environmental Research and Public Health" 20 [9], no. 5657.
- Raskin V. 1985. *Semantic Mechanisms of Humor*, D. Reidel, Dordrecht.
- Rodrigues D., Lopes D., Prada M., Thompson D. and Garrido M.V. 2017, *A Frown Emoji Can Be Worth a Thousand Words: Perceptions of Emoji Use in Text Messages Exchanged between Romantic Partners*, in "Telematics and Informatics" 34 [8], pp. 1532-1543.
- Rooksby M., Furuhashi T. and McLeod H.J. 2020, *Hikikomori: A Hidden Mental Health Need Following the COVID-19 pandemic*, in "World Psychiatry" 19 [3], pp. 399-400.
- Saito T. 1998 [trans. Angles J. 2013], *Hikikomori: Adolescence Without End*, University of Minnesota, Minneapolis.
- Sampietro A. 2019, *Emoji and Rapport Management in Spanish WhatsApp Chats*, in "Journal of Pragmatics" 143, pp. 109-120.
- Suwa M. and Suzuki K. 2013, *The Phenomenon of "Hikikomori" (Social Withdrawal) and the Socio-cultural Situation in Japan Today*, in "Journal of Psychopathology" 19, pp. 191-198.
- Tateno M., Teo A.R., Ukai W., Kanazawa J., Katsuki R., Kubo H. and Kato T.A. 2019, *Internet Addiction, Smartphone Addiction, and Hikikomori Trait in Japanese Young Adults: Social Isolation and Social Network*, in "Frontiers in Psychiatry" 10, 455.
- Togans L.J., Holtgraves T., Kwon G. and Morales Zelaya T.E. 2021, *Digitally Saving Face: An Experimental Investigation of Cross-cultural Differences in the Use of Emoticons and Emoji*, in "Journal of Pragmatics" 186, pp. 277-288.
- Widiana Y., Abdullah S.N.S. and Sumarlam S. 2024, *Phatic Emoji of Javanese Netizens in Social Media: A Cyberpragmatics Perspective on a Creative Communicative Strategy*, in "Creativity Studies" 17 [2], pp. 589-600.
- Yus F. 2011, *Cyberpragmatics: Internet-mediated Communication in Context*, John Benjamins, Amsterdam/Philadelphia.
- Yus F. 2014, *Not All Emoticons Are Created Equal*, in "Linguagem em (Dis)curso" 14 [3], pp. 511-529.

MULTIMODAL PRAGMATICS AT ITS BEST A Qualitative Analysis of ASMR Videos on TikTok

FABIO CIAMBELLA
SAPIENZA UNIVERSITY OF ROME

Abstract – This paper employs a multimodal pragmatic framework to analyse how creators of Autonomous Sensory Meridian Response (ASMR) videos on TikTok strategically use semiotic resources to improve digital intimacy and induce relaxation. While the sensory and psychological effects of ASMR are well-documented (Barratt, Davis 2015; Poerio *et al.* 2018), its linguistic and pragmatic mechanics remain underexplored (Battista 2025). This study addresses this gap by qualitatively analysing a corpus of popular TikTok ASMR videos to identify recurring pragmatic strategies. It examines how whispered speech acts – primarily directives and pseudo-rogatives – are fused with visual, aural, gestural, and spatial modes to create multimodal pragmatic alignments that blur the line between illocutionary intent and perlocutionary effect. The analysis further investigates whether and how these performances are perceived as polite (Lakoff 1983; Brown, Levinson 1987), balancing simulated personal care with subtle imposition to sustain engagement. Findings illustrate that TikTok ASMR is a sophisticated pragmatic genre where multisensorial communication is leveraged to perform care, build parasocial affiliation (Zappavigna 2019), and cater to contemporary wellness needs within the platform’s short-form format. This analysis expands pragmatic research into digital discourse, highlighting ASMR as a prime case study for how foundational linguistic concepts are reconfigured in multimodal, affectively charged online environments.

Keywords: ASMR; multimodal pragmatics; speech acts; digital intimacy; politeness; TikTok.

1. Introduction

Autonomous Sensory Meridian Response (ASMR) videos have evolved from a niche internet phenomenon into a mainstream digital genre, particularly on TikTok. ASMR refers to a tingling, static-like sensation that typically begins on the scalp and moves down the back of the neck and upper spine, which is triggered by specific auditory and visual stimuli. This sensation, often associated with feelings of calm and euphoria, is highly individualised and relies on what creators and viewers alike refer to as ‘triggers’. According to Barratt and Davis (2015, n.p.), “the stimulus that induces the pleasurable ASMR response is referred to as a ‘trigger’”, which commonly includes

“whispering (75%), personal attention (69%), crisp sounds (64%) and slow movements (53%)”. Similarly, Wang (2022) lists the following among the most common ASMR triggers: chewing sounds, tapping sounds, sounds of tools, sounds of nails, and whispering.

While the clinical and psychological effects of ASMR have been widely documented (see, among others, Marsden 2012; Taylor 2013; Barratt, Davis 2015), its linguistic and pragmatic mechanics remain underexplored (see Battista 2025). This paper addresses this gap by investigating how TikTok ASMR creators orchestrate multimodal resources to simulate intimacy and induce relaxation.

The study employs a multimodal discourse analysis framework to examine the concept of “digital intimacy” (Wang 2023) as it is constructed in a one-to-many mediated context. Following a review of ASMR and the specific affordances of the TikTok platform, the article outlines an analytical framework grounded in Austin’s (1962) and Searle’s (1969) Speech Act Theory (SAT) and (im)politeness frameworks (Lakoff 1983; Brown, Levinson 1987; Culpeper 1996 and ff.). The core of the study presents a qualitative analysis of recurring pragmatic strategies, including whispered directives and simulated personal attention. By analysing what I will define later as “multimodal pragmatic (mis)alignments”, the article demonstrates how creators leverage platform-specific constraints to blur the boundaries between illocutionary intent and perlocutionary effect, ultimately redefining the performance of care in digital spaces.

2. Literature review (...and beyond!): ASMR, multimodality, and pragmatics

2.1. The foundations of multimodal pragmatics

Multimodal pragmatics, a subfield that studies how meaning is constructed not just through language but through a range of semiotic resources (such as gaze, gesture, visual layout, and sound), offers the essential lens through which to investigate contemporary digital phenomena like ASMR. A major contribution to its theoretical grounding comes from social semiotics, especially the foundational works of Kress and van Leeuwen (2001, 2006) on multimodal discourse. They argue that each semiotic mode – whether linguistic, visual, audio/aural, or gestural – possesses distinct affordances and constraints for meaning-making. Images, for instance, can convey complex spatial and relational information at a glance, while speech unfolds sequentially over time, allowing for nuance, elaboration, and real-time negotiation of meaning. Van Leeuwen’s (1999) earlier research on sound

highlighted how vocal qualities such as pitch, rhythm, and loudness communicate subtle social meanings that language alone might obscure.

Empirical research in multimodal conversation analysis (CA) has further shown how core pragmatic phenomena are enacted through the dynamic orchestration of speech and embodied conduct. Drew (1992) illustrated how lawyers and witnesses in courtroom discourse strategically combine verbal argumentation with gesture, bodily orientation, and gaze to construct persuasive and credible narratives. Similarly, Goodwin (2007) demonstrated that speakers use gaze not only to allocate turns but to monitor hearer engagement, while hearers use nods, shifts in posture, or facial expressions to display alignment or resistance. Streeck's (2009) work provides compelling evidence that hand gestures – whether deictic, depictive, or emblematic – do not merely accompany speech but shape the pragmatic force of utterances.

The study of (im)politeness and stance-taking has likewise expanded to account for the multimodal modulation of pragmatic force. Haugh (2013) highlights how a compliment delivered with exaggerated intonation, eye-rolling, or ironic facial expressions may be reinterpreted as sarcasm, illustrating how semiotic modes can generate pragmatic dissonance. Du Bois's (2007) “stance triangle” framework emphasises that speakers align or disalign not only through propositional content but also through prosody, gaze, and gesture. Even fundamental pragmatic mechanisms such as deixis rely on the coordination of verbal and non-verbal resources (Hanks 1992; Mondada 2012).

The rise of digital media has introduced novel configurations of multimodal interaction. Social platforms like TikTok, Instagram, and YouTube embed multimodality into their very design, constraining and enabling meaning-making through technical affordances such as video length, editing tools, filters, and interactive features (Androutsopoulos 2014). Lee's (2025) study of TikTok's “duet” feature shows how users construct layered, intertextual performances that challenge traditional pragmatic concepts such as deixis and reference, requiring users to navigate fragmented spatio-temporal frames.

2.2. ASMR as a case study in “digital intimacy”

Against this theoretical backdrop, Autonomous Sensory Meridian Response (ASMR) emerges as a paradigm case for multimodal pragmatic analysis. As Wang (2022, n.p.) notes, “ASMR, born with visual mode and audio mode”, is inherently multimodal and grounded in what she refers to as “visual and audio grammars”. Drawing on foundational work by Halliday (1985) and Kress and van Leeuwen (2006), Wang adopts the notion of visual grammar to examine how images function in ASMR videos. For instance, a “demand act”

typically establishes “an imaginary relation with the viewer since they address her/him directly” (Stoian 2015, p. 26). In ASMR contexts, the “represented symbols” are often the content creators themselves, who “gaze and talk with [the] audience to demand [their] attention and assume the presence of audience which invites [them] to be involved in imaginary relations” (Wang 2022, n.p.).

In addition to visual grammar, ASMR content relies heavily on “audio grammar”, as conceptualised by van Leeuwen (1999). Audio elements such as whispering, tapping, and soft ambient noise are fundamental to the immersive, affective experiences ASMR videos aim to produce. Wang (2022) underscores the link between auditory intimacy and soft vocal delivery: “intimate distance is always related to whispering or maximally soft sounds”. The role of visual framing further supports this idea. A “close-up shot, focusing on heads and shoulders or participants’ bodies above waists, suggests intimate or personal relation” (Wang 2022, n.p.). Barratt and Davis (2015, n.p.) similarly observe that “many of these videos depict role play situations, in which the viewer is placed in a position of ‘close proximity’ to another person to be cared for in some manner”.

These strategies fall under what Wang (2023) refers to as “digital intimacy” – a phenomenon whereby creators simulate closeness and personal connection with viewers despite the inherent asymmetry and physical distance of digital platforms. This digital intimacy is often supported by what Zappavigna (2019, 2021) terms “ambient affiliation”, or the affective bonding between users within shared digital environments. In ASMR, ambient affiliation is built within the very structure and delivery of the video content. Creators frequently speak in soothing tones, employ second-person pronouns, and simulate direct address to evoke a personal relationship. These are not just aesthetic choices but pragmatic ones, embedded in the deliberate construction of an affective stance (Bucholtz and Hall 2005).

In ASMR, the typical pragmatic distinction between the illocutionary act and the perlocutionary goal is fundamentally blurred. This occurs because the function of the speech is often identical to its intended effect. When an ASMR artist whispers a phrase like, “It’s okay, relax”, the illocutionary function of soothing reassurance is the direct mechanism for triggering the perlocutionary goal of physiological relaxation. The emphasis shifts from the meaning of the words to the auditory texture, making the act of speaking inseparable from the desired effect it immediately produces.

The relevance of ASMR to linguistic study lies in its demonstration of how language functions within complex, affectively charged, and digitally mediated environments. As Zappavigna (2021) notes, ASMR is one of the clearest examples of how social media practices generate new genres that challenge existing linguistic models.

2.3. Positioning the study: Multimodal pragmatic alignment and misalignment

In this article, I draw on the classic pragmatic frameworks reviewed – Austin¹ and Searle’s Speech Act Theory (SAT), Grice’s Cooperative Principle (CP), Lakoff’s Politeness Principle (PP), and Brown and Levinson’s Politeness Theory – to analyse the contemporary, highly multimodal text of ASMR videos on TikTok. By examining how linguistic, visual, audio/aural, gestural, and spatial modes interact, I will explore two key phenomena derived from this theoretical synthesis. First, *multimodal pragmatic misalignments*, where conflicting semiotic modes generate divergent illocutionary or perlocutionary effects, producing irony or ambiguity. Second, and more central to ASMR, *multimodal pragmatic alignments*, where modes synergistically align to reinforce the intended pragmatic force of a communicative act, creating heightened affective resonance.

This approach demonstrates how ASMR creators mobilise multimodality to either disrupt or, more commonly, intensify pragmatic effects. Ultimately, this study situates ASMR as an instructive site for testing how foundational pragmatic notions are complicated or enhanced when speech interacts with embodied, sensory, and technological modes in online interaction, providing new insights into the co-construction of meaning and intimacy across sensory channels in digital discourse.

3. A multimodal pragmatic framework for ASMR videos

My qualitative analysis was conducted on a small, self-compiled sample corpus of 100 ASMR videos on TikTok. These videos were not explicitly selected, so any bias in the sample is the product of the algorithm and not my decision-making. Crucially, only videos in which content creators whisper complete utterances directly to their viewers² – rather than relying solely on mouth sounds – were included in the corpus. This choice foregrounds the role of language in ASMR content that combines spoken interaction with sensory triggers.

¹ Nevertheless, Austin was probably the first scholar who recognised the multimodal value of speech acts, stating that “the uttering of the sentence is, or is a part of, the doing of an action, which again would not *normally* be described as, or as ‘just’, saying something” (1962, p. 5; emphases in the original).

² ASMR videos can also involve a ‘third party’ who receives care and attention from the content creator, simulating what they would do for their follower. Generally, the third party is a person or an animal – usually a dog – who is massaged on the head. This kind of video has not been included in the corpus.

The analysis examined each video through the lens of the multimodal framework proposed by the New London Group,³ which identifies five key modes, as hinted at earlier:

- 1) Linguistic: The use of language – specifically, fully formed whispered utterances – is central to this corpus. While some ASMR videos rely exclusively on non-verbal mouth sounds to induce relaxation, the selected sample highlights the communicative dimension of whispered speech as an integral part of the viewer's sensory experience.
- 2) Visual: Visual elements, including lighting, colour schemes, and subtle visual effects, play a crucial role in shaping the sensations elicited by ASMR. Soft, calming lights and deliberate colour choices contribute to the immersive, soothing atmosphere that characterizes these videos (see Figs. 1, 2 and 3).
- 3) Audio/Aural: A wide range of mouth sounds (such as whispers, kissing noises, fluttering, chewing, swirling, lip smacking, and tongue clicking) are employed alongside other auditory triggers. These include sounds generated by objects like brushes gently tapping the microphone, glass cruets, or fingernails softly scratching various surfaces (see Fig. 1). Such sounds are often regarded as the primary stimuli that produce the tingling and relaxing effects for which ASMR is known.
- 4) Gestural: The gestural mode encompasses deliberate hand movements, which are carefully orchestrated to capture and guide the viewer's visual attention (see Fig. 2). Quick finger flutters, slow tracing motions, or simulated touches enhance the intimate, personal nature of the ASMR experience.
- 5) Spatial: ASMR videos typically position the content creator in the immediate foreground – except for those that involve a 'third party', which however are not included in the corpus analyzed here – with the background kept intentionally simple – often dark or softly illuminated with calming hues to reduce distractions and foster a sense of tranquillity (see Figs. 1 and 2). Many creators break the fourth wall through gestures that simulate direct interaction with the viewer, such as pretending to touch the viewer's face or hair, performing light

³ The New London Group, a collective of scholars including Norman Fairclough, the father of Critical Discourse Analysis, and Gunther Kress, one of the milestones of multimodal studies, among others, pioneered the concept of multiliteracies and multimodal discourse analysis in their seminal 1996 article, *A Pedagogy of Multiliteracies: Designing Social Futures*. They argued that communication extends beyond language alone, emphasizing the interplay of linguistic, visual, audio/aural, gestural, and spatial modes in meaning-making. Their framework has been foundational in shifting literacy studies toward a broader understanding of how multiple semiotic systems interact in digital, visual, and embodied texts – an approach central to this article's analysis of multimodal pragmatics in ASMR videos.

makeup, or offering gentle massages, thereby enhancing the immersive, one-on-one dynamic that defines this genre (see Fig. 3).



Figure 1

An example of an auditory trigger realised by moving a brush around the microphone.

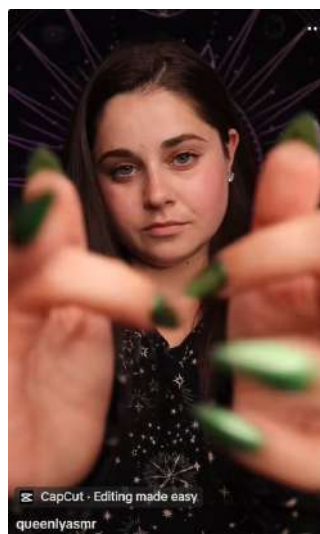


Figure 2

An example of quick hand movements to capture and guide the viewer's visual attention.



Figure 3

An example of how ASMR content creators try to interact with their followers, breaking the fourth wall (in this case by pretending to ask viewers if they can taste some of the food they are eating).

To pragmatically analyse the verbal language in these videos, a reference to Speech Act Theory (SAT), founded by J.L. Austin (1962) and developed by John Searle (1969), is essential. This revolutionary theory proposed that saying something is always doing something. Austin distinguished three acts within an utterance: the locutionary act (the production of sounds with meaning), the illocutionary act (the communicative force or intention, such as ordering, promising, inviting), and the perlocutionary act (the effect on the listener, such as convincing, calming, frightening). Searle systematised the theory by classifying illocutionary acts like declaratives, commissives, expressives, directives, and assertives. SAT thus provides the conceptual tools to identify not only what is said in an ASMR video, but what the creator is doing through those words: are they issuing a gentle command (“relax”), offering a service (“let me take care of you”), or expressing empathy (“I know you’re tired”)? This focus on the speaker’s action is the fundamental first step in understanding the pragmatic mechanics of the digital intimacy constructed in ASMR.

However, applying classic SAT to an inherently multimodal phenomenon like ASMR requires adaptation. In the analysed videos, the illocutionary force of an invitation or an expression of care is rarely entrusted to words alone. On the contrary, it is co-constructed and amplified by the

convergence of all five semiotic modes. A whisper (linguistic/aural) devoid of close-up gaze (visual) and slow gestures (gestural) would lose much of its calming force and performative genuineness. Therefore, the following analysis will not merely catalogue linguistic acts but will explore how these fuse with other modes to create what I define as multimodal pragmatic alignments, where meaning and pragmatic force emerge from the perfect synergy of different channels. It is precisely this synergy that generates the desired perlocutionary effect: the tingling or relaxation in the viewer.

Besides examining which speech acts ASMR videos on TikTok most commonly rely on, another interesting and non-trivial research question is whether such videos, in which content creators invite their followers to relax, may or may not be considered polite.

Using H. Paul Grice's CP (1975), Robin T. Lakoff (1973) developed her Politeness Principle (PP). Grice's CP focuses on the idea that communication is a cooperative activity where speakers and listeners work together to achieve mutual understanding by following maxims of quantity, quality, relation, and manner – the famous four Gricean maxims. Lakoff's PP builds on this, suggesting that speakers also consider social factors and attempt to minimise social friction through strategies like being indirect, showing deference, and being friendly. Therefore, beyond simply being cooperative, speakers also aim to be polite, which involves considering social dynamics and minimising potential social friction. For Lakoff, pragmatic competence is reached if two 'rules' are respected: 'Be clear', that is, follow Grice's CP, and 'be polite', which is further divided into three subrules: 'Don't impose' (R1), 'Give options' (R2), and 'Be friendly' (R3), as schematised by Watts (2003, p. 60) and re-adapted by Flowerdew (2013, p. 107), and shown in Fig. 4:

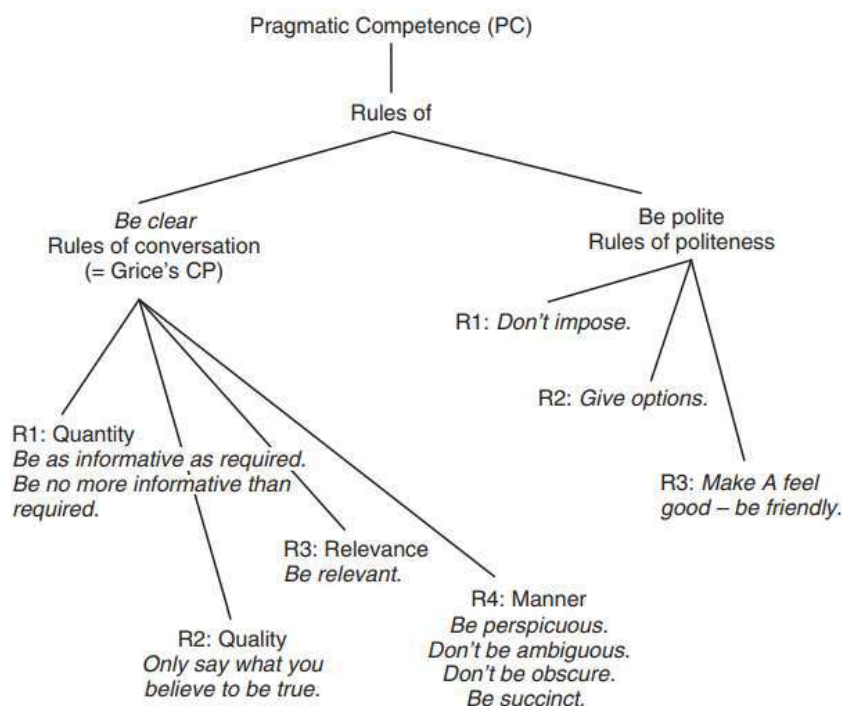


Figure 4

Lakoff's scheme of pragmatic competence (Watts 2003, p. 60; Flowerdew 2013, p. 107).

Lakoff's principle offers an effective lens for interrogating the very nature of ASMR interaction. While on one hand these videos scrupulously respect Grice's CP (the instructions are clear, the role-play is coherent), on the other, they pose an interesting challenge to the sub-rules of politeness. The invitation to "not scroll away" or to "relax" is, in essence, a form of imposition (R1) on the user's freedom, albeit cloaked in friendliness. Similarly, while the option to close the video technically exists (R2), the entire multimodal apparatus is designed to make that option increasingly unlikely, creating an affective constraint. This tension between friendly imposition and free choice leads us directly to a more articulated theory of politeness that accounts for individuals' social 'face': the theory of Brown and Levinson (1987). Their model, centred on the Goffmanian (1967) notion of 'face' (public self-image) and the strategies to mitigate threats to it (Face-Threatening Acts, FTA), proves particularly suited to decoding the sophisticated performances of intimacy in ASMR.

In particular, I interpret ASMR videos as a fertile ground for observing positive politeness strategies, which Brown and Levinson define as "redress directed to the addressee's positive face, his perennial desire that his wants (or the actions/acquisitions/values resulting from them) should be thought of as desirable" (1987, p. 101). Central to positive politeness is the idea that it

is not necessarily redressive of the particular face want infringed by the FTA; that is, whereas in negative politeness the sphere of relevant redress is

restricted to the imposition itself, in positive politeness the sphere of redress is widened to the appreciation of alter's wants in general or to the expression of similarity between ego's and alter's wants. [...] [T]he linguistic realisations of positive politeness are in many respects simply representative of the normal linguistic behaviour between intimates, where interest and approval of each other's personality, presuppositions indicating shared wants and shared knowledge, implicit claims to reciprocity of obligations or to reflexivity of wants, etc. are routinely exchanged. (p. 101)

And again,

Perhaps the only feature that distinguishes positive-politeness redress from normal everyday intimate language behaviour is an element of exaggeration; this serves as a marker of the face-redress aspect of positive-politeness expression, by indicating that even if S can't with total sincerity say 'I want your wants', he can at least sincerely indicate 'I want your positive face to be satisfied.' Thus the element of insincerity in exaggerated expressions of approval or interest [...] is compensated for by the implication that the speaker really sincerely wants Mrs B's positive face to be enhanced. (pp. 101, 103)

Positive politeness strategies are exemplified in Fig. 5 below:

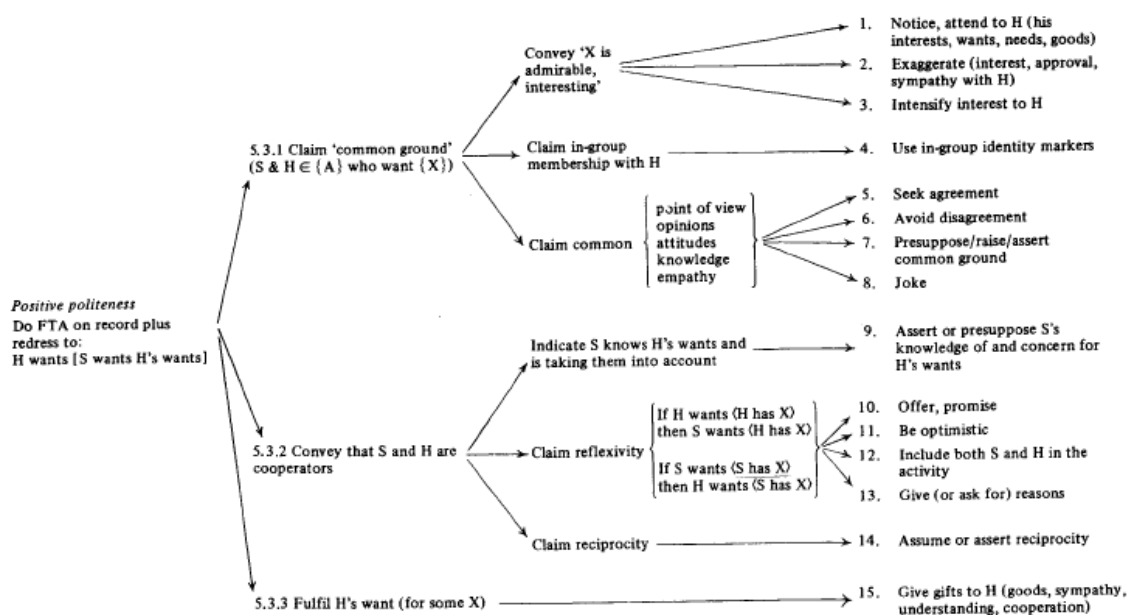


Figure 5

Brown and Levinson's chart of positive politeness strategies (1987, p. 102).

The complex framework outlined above provides the foundation for the analysis that follows. The methodology has been established – a multimodal analysis of selected ASMR videos, examining five distinct semiotic modes – and the key pragmatic concepts have been introduced: SAT to identify what creators do with words, and Lakoff's and Brown and Levinson's models to

interrogate the polite or impolite nature of those acts. Having adapted these linguistic theories to account for the multimodal reality of ASMR, the analytical section can now proceed. It will apply this specific toolkit to the corpus, investigating the central phenomena this framework is designed to illuminate: how multimodal pragmatic alignments create intimate, perlocutionary effects, and how the performance of positive politeness navigates the tension between simulated care and strategic engagement in digital spaces.

4. Analysis

4.1. *Multimodal speech acts*

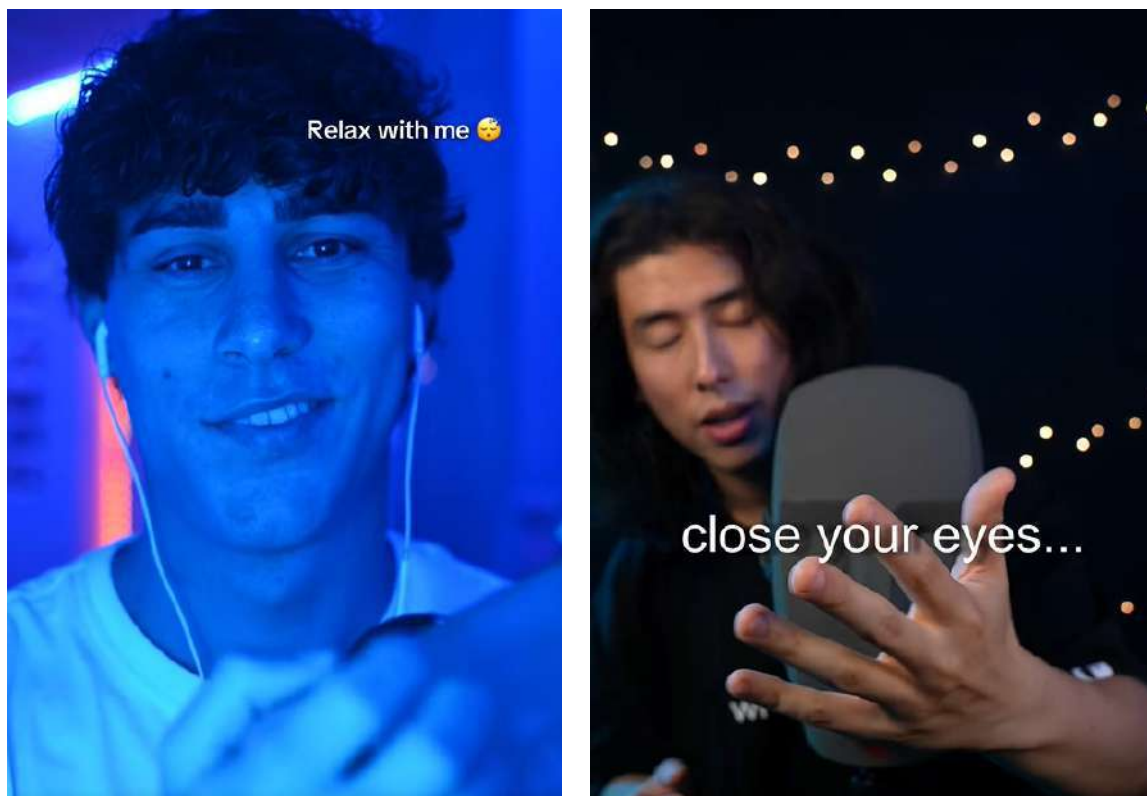
Among the various speech acts involved in the multimodal framework⁴ of ASMR videos, directives (Searle 1969) and rogatives (Leech 1983) are arguably the most intriguing to analyse. Directives, i.e., attempts by the speaker to prompt the hearer to perform an action, are generally associated with commands, requests, orders, and invitations, and typically employ the imperative mood. They form the very essence of ASMR videos, which are essentially continuous invitations by the speaker – the content creator – for the hearer – the viewer or follower – to sleep, relax, or focus.

I argue that the underlying principle guiding all the strategies examined here and in the following sections is what I have labelled as multimodal pragmatic alignment, wherein multiple semiotic codes work together to reinforce both one another and the illocutionary force of the multimodal utterances under consideration. In the specific case of directives, the invitation to relax or sleep is realised through the interplay of various modes. The linguistic code is, of course, the most explicit, exemplified by imperatives such as “Relax”, “Close your eyes”, “Stay still”, or “Fall asleep”. The audio/aural mode complements this through characteristic whispering and the elongation of final vowel sounds, which are often exaggerated and sustained in an unnatural but soothing way (e.g., *rɪ'læ:ks, *aɪ:z, *stɪ:l).

Moreover, directiveness is visually and spatially reinforced by hand movements that complete the invitation to unwind and drift off to sleep (Figs. 6 and 7). These gestures often include slow, deliberate motions, such as pointing toward the camera to simulate head massages or brushing movements, thereby breaking the fourth wall and enhancing the sense of intimacy between speaker and hearer. The gestural mode, in tandem with soft

⁴ It is worth noting that van Leeuwen (2004) defines multimodal speech acts as “communicative acts” (see also Chałupnik, Brookes 2022, p. 310). Nevertheless, since I want to emphasize the specifically multimodal dimension of speech acts in ASMR videos, I will not adopt this label.

lighting (visual mode) and ambient visuals, contributes to creating an immersive atmosphere that heightens the directive force.



Figures 6 and 7

Examples of directives aimed to invite the viewer to relax or sleep.

Taken together, these multimodal elements function harmoniously to create a persuasive and comforting environment, illustrating how ASMR content relies on the convergence of linguistic, aural, visual, gestural, and spatial codes to achieve its intended perlocutionary effects. In this sense, ASMR multimodal acts can be seen as what Austin (1962) identifies as acts where the perlocutionary dimension is predominant. While traditional speech act theory often separates intent from effect, Austin allows for certain acts – such as persuading or boasting – where the act itself is heavily geared toward, or even defined by, its perlocutionary outcome. In ASMR, the multimodal directive to relax is not merely a request for a future state but a perlocutionary-heavy performance where the acoustic and visual textures are the immediate instruments of the viewer's physiological response.

Rogatives, a speech act type added by Leech (1983, p. 206) to Searle's taxonomy, are utterances designed to elicit information from the listener, typically in the form of questions. In the context of ASMR, however, these acts often transcend their literal function of seeking information. The ASMR videos in the corpus frequently deploy rhetorical and pragmatic strategies that simulate dialogue through questions – what some linguists refer to as staged

or fictional dialogue or pseudo-questions, and which I will term pseudo-rogative speech acts. These pseudo-rogatives serve to enhance the illusion of genuine interpersonal interaction, which is central to the immersive and intimate experience of ASMR content.

Such pseudo-rogatives commonly take the form of hypophoras, that is, questions posed and immediately answered by the speaker, as well as proleptic questions, which anticipate or preempt the listener's imagined response (e.g., "You want some French tip nail? Okay, perfect", "Are you here for the hair care? Okay, I've got different brushes, clips, and hairsprays", "I was wondering if I could have you both for a couple of photos, okay? Yeah, you look great". See Figs. 8 and 9). These forms enable creators to shape and control the discourse in ways that subtly guide the viewer's interpretation while sustaining the illusion of a mutual, reciprocal exchange.



Figure 8

A pseudo-rogative realised as hypophora: the creator stages a choice ("Which tickle? 1, 2, or 3?") but then unilaterally performs all three options, simulating a dialogue to enhance personal attention while maintaining full control of the script.



Figure 9

A pseudo-rogative realised as a proleptic question. The creator anticipates the viewer's response by asking "is my asmr good" and immediately performing them. Audience interaction is deferred to the comments section after the video's performance, highlighting the asynchronous, staged nature of digital intimacy.

Moreover, by employing pseudo-rogatives, ASMR creators blur the boundary between monologue and dialogue, crafting a highly personalised, parasocial

interaction that makes the viewer feel directly addressed and cared for. This sense of imagined conversational intimacy not only strengthens the bond between content creator and audience but also amplifies the soothing and comforting qualities that define the genre. In this way, pseudo-rogative speech acts function as a key component of the multimodal framework of ASMR, working in tandem with visual, auditory, and gestural cues to construct an environment where viewers feel acknowledged, attended to, and gently guided into states of relaxation or sleep.

4.2. *Is ASMR polite? Multimodal politeness between friendliness and imposition*

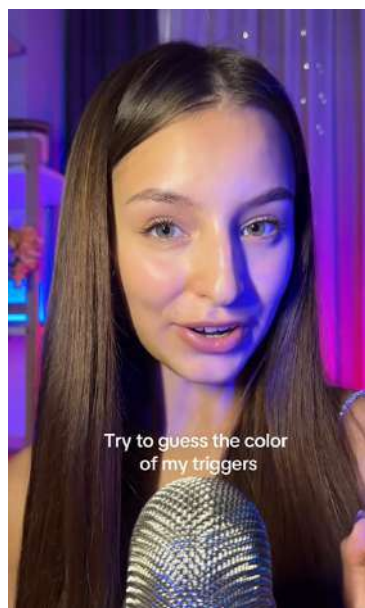
Beyond examining speech acts, it is crucial to investigate whether ASMR videos can be considered ‘polite’. As seen in Section 3, pragmatic competence, as defined by Lakoff (1973), relies on two rules: “Be clear” (Gricean Cooperation) and “Be polite”. While the former is generally respected in ASMR – as creators provide relevant and clear sensory triggers – the latter presents a fascinating tension. As stated above, Lakoff’s Politeness Principle (PP) is subdivided into three sub-rules: “Don’t impose” (R1), “Give options” (R2), and “Be friendly” (R3) (Watts 2003).

On the one hand, Grice’s CP, equivalent to Lakoff’s ‘Be clear’ rule, appears to be respected in all the ASMR videos in the analysed corpus. The speaker provides sufficient information for their intended purposes (maxim of quantity), says only what they believe to be true (maxim of quality) – although they sometimes adopt a persona (e.g., a doctor, a beautician) thus performing what is commonly known as roleplay – ensures that gestures and words are relevant to the pseudo-dialogues enacted (maxim of relevance), and generally avoids ambiguity (maxim of manner), often relying on aligned multimodal pragmatic techniques.

On the other hand, Lakoff’s PP is, in my view, sometimes neglected in the analysed corpus. While content creators consistently aim to be friendly, thereby fulfilling R3 (‘Be friendly’), R1 (‘Don’t impose’) and 2 (‘Give options’) are not always respected; a tendency also observed in other forms of media discourse (see, among others, Chandra 2021; Sudaryat *et al.* 2020). Although they often redress and mitigate their face-threatening acts (FTAs) – as will be discussed later with reference to Brown and Levinson’s face-based model of politeness, especially positive politeness strategies – ASMR video creators actively exploit the potential of multimodal pragmatic alignment to reinforce the strength of their invitations to relax, sleep, or remain engaged. In doing so, they effectively impose the illocutionary force of their utterances on their followers (thereby violating R1) and leave little room for genuine choices (violating R2).

This impositive dimension is particularly evident in videos that open

with explicit directives such as “Hey, you, don’t scroll down”, or other semantically and pragmatically similar utterances that seek to suspend the user’s freedom of action (see Figs. 10 and 11). In these cases, the linguistic code operates as an overt injunction, while the accompanying modes – auditory cues like soothing whispers, repetitive tapping, or ambient sounds; visual elements such as calming gestures or direct gaze; and spatial framing designed to simulate intimacy – work together to foster a sense of comfort and compliance.



Figures 10 and 11
Examples of directives as violations of Lakoff’s R1 and R2.

Such tension becomes even more pronounced in those (albeit fewer) videos where content creators paradoxically challenge or ‘threaten’ their followers not to fall asleep with imperatives like “Try not to sleep”. Here, the linguistic mode explicitly contradicts the relaxing sensory cues produced by audio, visual, gestural, and spatial modes, resulting in what can be described as multimodal pragmatic misalignments. Rather than simply inviting relaxation, these videos create a playful yet coercive dynamic, engaging viewers in a pseudo-competition that paradoxically amplifies their attentiveness and emotional investment.

In this sense, while the illusion of user agency persists, since viewers can technically choose to disengage at any moment, the performative structure of such videos subtly constrains that freedom by blending friendliness with soft imposition. It goes without saying that in contemporary digital discourse, R2 of Lakoff’s PP is almost always respected, since TikTok users, like social media users in general, implicitly retain the option to scroll away at any time. However, the pragmatic force of these multimodal invitations often works precisely to delay or undermine that very choice,

sustaining engagement through a complex interplay of alignment and misalignment among modes.

The analysis carried out so far has gradually shifted our focus from understanding ASMR videos solely as social media content designed to induce a sense of relaxation in viewers to viewing them as digital products that strategically employ specific multimodal pragmatic techniques to increase their creators' follower counts. After all, this dual objective lies at the heart of most digital and social media content today, and there is nothing contradictory about acknowledging that an emphasis on both the viewers' mental well-being and the creators' audience growth can coexist harmoniously.

These final paragraphs consider TikTok ASMR videos – at least those included in the selected corpus – as an exemplary site for analysing politeness from the face-based perspective developed by Brown and Levinson (1987) and summarised in Section 3. As Brown and Levinson point out, positive politeness strategies are linguistically very similar, if not identical, to the kind of language typically used among intimates. The key difference, however, lies in the element of exaggeration and, consequently, a degree of insincerity that characterises positive politeness when used outside genuinely close relationships. I argue that this distinction between intimate (sincere) language and the strategic use of positive politeness has clear parallels with the dual nature of ASMR videos, which, as discussed above, can be understood both as content produced for the viewers' mental well-being and as content designed to attract and retain followers. If one adopts the first perspective, ASMR videos draw on multimodal resources that mimic the comforting, familiar communication typical of intimate interactions. In contrast, if one adopts the second perspective, the same techniques appear as calculated positive politeness strategies that risk seeming exaggerated or insincere.

This second interpretation becomes especially evident when content creators employ visual and gestural modes that simulate close, personal engagement. For example, they often hold up brightly coloured objects or lights, asking viewers to follow them with their eyes or focus intently, and then respond with emphatic praise such as “Perfect!”, “Very good!”, or “You’re amazing!” regardless of whether they can verify the viewer’s compliance. Such formulaic expressions echo Brown and Levinson’s positive politeness strategy no. 2⁵ (exaggerating interest, approval, or sympathy), reinforcing the illusion of intimacy while primarily serving to sustain attention and engagement.

⁵ In the literature, this is typically abbreviated as PP2, but I will use the extended form to avoid confusion with Lakoff’s Politeness Principle discussed above.

In this sense, the praise acts less as genuine feedback than as a performative device that blurs the line between authentic care and strategic audience management. This ambiguity invites viewers to negotiate for themselves whether the interaction feels truly supportive or merely performative – a tension that lies at the heart of how ASMR content navigates the pragmatics of positive politeness in digital contexts.

Other positive politeness strategies are also employed by different content creators, depending on their personal delivery style and the specific pseudo-intimacy they aim to construct. For instance, video makers who open their videos with proleptic questions such as “Sorry? Do you want me to help you with your makeup? Okay” (or similar utterances), demonstrate the use of strategies no. 1 (‘Notice, attend to H’) and no. 9 (‘Assert or presuppose speaker’s knowledge of and concern for H’s wants’). By pre-emptively anticipating the viewer’s desire, these creators simulate a conversational exchange that foregrounds attentiveness and care, reinforcing the impression of a shared interpersonal space.

Similarly, some ASMR performers repeat the same gesture for themselves first and then toward the camera – for example, pretending to taste food with a spoon and then offering it to the viewer – which illustrates strategy no. 14 (‘Assume or assert reciprocity’). This act of mirroring not only implies a shared experience but also constructs a symbolic bond between creator and viewer, suggesting a mutual involvement in the interaction. Such reciprocal gestures enhance the illusion of co-presence and equal participation, even though the communication remains one-sided.

These examples highlight how ASMR video makers skilfully adapt a range of positive politeness strategies to shape an affective atmosphere that combines attentiveness, care, and shared ritual. While these techniques draw heavily on familiar face-to-face conversational norms, their repetition and stylisation within a scripted, multimodal performance underscore the tension between authentic relational cues and calculated audience engagement. In this way, the videos sustain a delicate balance between seeming spontaneous and remaining highly orchestrated, a dynamic that is central to understanding their pragmatic complexity.

5. Conclusion

This study has employed a multimodal pragmatic framework to analyse how intimacy and relaxation are orchestrated within TikTok ASMR videos. By examining the interplay of linguistic, audio, visual, gestural, and spatial modes, the analysis demonstrates that the genre’s effectiveness hinges on multimodal pragmatic alignments, where semiotic resources synergistically reinforce a unified illocutionary and perlocutionary force. Central speech

acts, such as directives and pseudo-rogatives, are realised not merely through whispered language but through a coordinated sensory ensemble: soft prosody, close-up visuals, simulated touch, and an intimate spatial framing that collectively construct an immersive parasocial encounter.

The investigation further reveals a nuanced tension within the pragmatics of ASMR. While these videos meticulously follow Gricean cooperative maxims to provide clear, engaging triggers, they often navigate Lakoff's PP ambivalently. Creators masterfully employ positive politeness strategies – attending to the viewer's wants, exaggerating approval, and simulating reciprocity – to foster a digital intimacy that feels personally attentive. However, this very simulation can slip into a soft imposition, where the aligned multimodal cues gently constrain the viewer's agency, prioritising sustained engagement and therapeutic effect over genuine optionality. Thus, ASMR on TikTok emerges as a complex pragmatic performance where the boundaries between authentic care and strategic design, and between illocutionary intent and perlocutionary effect, are productively blurred.

Ultimately, this analysis underscores the value of applying classic pragmatic theories to contemporary, multimodal digital genres. ASMR is not merely a relaxing pastime but a sophisticated discourse that redefines notions of intimacy, politeness, and speech acts in a one-to-many mediated context. It showcases how digital platforms like TikTok enable new forms of affective communication, where the orchestration of multiple modes becomes the primary mechanism for building ambient affiliation and performing care. Future research could productively extend this framework to examine longitudinal viewer responses, cross-cultural variations in ASMR triggers, or the ethical dimensions of parasocial relationships cultivated through such meticulously engineered pragmatic alignments.

Bionote: Fabio Ciambella is a Researcher of English 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).

Author's address: fabio.ciambella@uniroma1.it

Acknowledgements: My sincere thanks go to the reviewers for their insightful feedback, which greatly improved the coherence of this article.

References

- Androutsopoulos J. (ed.) 2014, *Mediatization and Sociolinguistic Change*, Walter de Gruyter, Berlin/Boston.
- Austin J.L. 1962, *How to Do Things with Words*, Oxford University Press, Oxford.
- Barratt E.L. and Davis N.J. 2015, *Autonomous Sensory Meridian Response (ASMR): A Flow-like Mental State*, in "PeerJ" 3, no. e851.
- Battista A. 2025, 'Delicious Deliciousness', 'Therapeutic' Peeling, and the Effect of Sensory Language and of ASMR/Satisfying Triggers on Authenticity: A Corpus-based Multimodal Analysis of Tasty's Videos, in D'Amore M., Zago R., Bugliani P., Polatti A. and Vignozzi G. (ed.), *English Studies in Italy: New Directions and Perspectives*, Carocci, Roma, pp. 23-40.
- Bhandari A. and Bim S. 2020, *TikTok and the 'Algorithmized Self': A New Model of Online Interaction*, in *Selected Papers of #AoIR2020: The 21st Annual Conference of the Association of Internet Researchers*. <https://spir.aoir.org/ojs/index.php/spir/article/view/11172/9856> (21.06.2025).
- Brown P. and Levinson S.C. 1987, *Politeness: Some Universals in Language Usage*, Cambridge University Press, Cambridge.
- Bucholtz M. and Hall K. 2005, *Identity and Interaction: A Sociocultural Linguistic Approach*, in "Discourse Studies" 7 [4-5], pp. 585-614.
- Chałupnik M. and Brookes G. 2022, *Discursive Acts of Resistance: A Multimodal Critical Discourse Analysis of All-Poland Women's Strike's Social Media*, in "Gender and Language" 16 [3], pp. 308-333.
- Chandra O.H. 2021, *Politeness in the Use of Language in Social Media*, in "E3S Web of Conferences" 317, no. 02027.
- Drew P. 1992, *Contested Evidence in Courtroom Cross-examination: The Case of a Trial for Rape*, in Drew P. and Heritage J. (ed.), *Talk at Work: Interaction in Institutional Settings*, Cambridge University Press, Cambridge, pp. 470-520.
- Du Bois J.W. 2007, *The Stance Triangle*, in Englebretson R. (ed.), *Stancetaking in Discourse: Subjectivity, Evaluation, Interaction*, John Benjamins, Amsterdam/Philadelphia, pp. 139-182.
- Engelbregt H.J., Brinkman K., van Geest C.C.E., Irmischer M. and Deijen J.B. 2022, *The Effects of Autonomous Sensory Meridian Response (ASMR) on Mood, Attention, Heart Rate, Skin Conductance and EEG in Healthy Young Adults*, in "Experimental Brain Research" 240 [6], pp. 1727-1742.
- Flowerdew J. 2013, *Discourse in English Language Education*, Routledge, London/New York.
- Goffman E. 1967, *Interaction Ritual: Essays on Face-to-Face Behavior*, AldineTransaction, New Brunswick (NJ)/London.
- Goodwin C. 2000, *Action and Embodiment within Situated Human Interaction*, in "Journal of Pragmatics" 32 [10], pp. 1489-1522.
- Goodwin C. 2007, *Participation, Stance, and Affect in the Organization of Activities*, in "Discourse and Society" 18, pp. 53-73.
- Grice H.P. 1975, *Logic and Conversation*, in Davidson D. (ed.), *The Logic of Grammar*, Dickenson Publishing Company, Encino (CA), pp. 64-75.
- Halliday M.A.K. 1985, *An Introduction to Functional Grammar*, Edward Arnold, London.
- Hanks W.F. 1992, *The Indexical Ground of Deictic Reference*, in Duranti A. and Goodwin C. (ed.), *Rethinking Context: Language as an Interactive Phenomenon*, Cambridge University Press, Cambridge, pp. 43-76.
- Haugh M. 2013, *Im/politeness, Social Practice and the Participation Order*, in "Journal of Pragmatics" 58, pp. 52-72.

- Kádár D.Z. and Haugh M. 2013, *Understanding Politeness*, Cambridge University Press, Cambridge.
- Kress G., van Leeuwen T. 2001, *Multimodal Discourse: The Modes and Media of Contemporary Communication*, Arnold Publishers, London.
- Kress G. and van Leeuwen T. 2006, *Reading Images: The Grammar of Visual Design*, Routledge, London/New York.
- Krueger J. 2009, *Enacting Musical Experience*, in “Journal of Consciousness Studies” 16 [2-3], pp. 98-123.
- 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*, Department of Linguistics, University of Chicago, Chicago, pp. 292-305.
- Lee Y.-J. 2025, *Language Learning with TikTok's 'Duet': A Spatial Perspective on Digital Technology and Language Learning*, in “Innovation in Language Learning and Teaching”, pp. 1-23.
- Leech G.N. 1983, *Principles of Pragmatics*, Longman, London/New York.
- Marsden R. 2012, “*Maria spends 20 minutes folding towels*”: *Why Millions Are Mesmerised by ASMR Videos*, in “The Independent”, 20.07.2012. <https://web.archive.org/web/20200527013456/https://www.independent.co.uk/life-style/gadgets-and-tech/features/maria-spends-20-minutes-folding-towels-why-millions-are-mesmerised-by-asmr-videos-7956866.html> (26.12.2025).
- Mondada L. 2012, *Deixis: An Integrated Interactional Multimodal Analysis*, in Bergmann P., Brenning J., Pfeiffer M. and Reber E. (ed.), *Prosody and Embodiment in Interactional Grammar*, De Gruyter, Berlin, pp. 173-206.
- Mondada L. 2016, *Challenges of Multimodality: Language and the Body in Social Interaction*, in “Journal of Sociolinguistics” 20 [3], pp. 336-366.
- Poerio G.L., Blakey E., Hostler T.J. and Veltri T. 2018, *More than a Feeling: Autonomous Sensory Meridian Response (ASMR) is Characterized by Reliable Changes in Affect and Physiology*, in “PLOS One” 13 [6], no. e0196645.
- Schellewald A. 2023, *Understanding the Popularity and Affordances of TikTok through User Experiences*, in “Media, Culture & Society” 45 [8], pp. 1568-1582.
- Searle J. 1969, *Speech Acts: An Essay in the Philosophy of Language*, Cambridge University Press, Cambridge.
- Stoian C. 2015, *Analysing Images: A Social Semiotic Perspective*, in “Scientific Bulletin of the Politehnica University of Timișoara: Transactions on Modern Languages” 14 [1], pp. 23-30.
- Streeck J. 2009, *Gesturecraft: The Manufacture of Meaning*, John Benjamins, Amsterdam/Philadelphia.
- Sudaryat Y., Widyastuti T. and Hernawan 2020, *Politeness on the Social Media: A Study on the Response of West Java City Information Instagram Account*, in “Advances in Social Science, Education and Humanities Research” 509, pp. 792-798.
- Taylor S. 2013, ‘*Head Orgasms*’, *Meditation and Near Death Experiences*, in “The Guardian”, 09.10.2013. <http://www.theguardian.com/science/brain-flapping/2013/oct/09/head-orgasms-meditation-near-death-experiences> (19.06.2025).
- Taylor V. 2014, *Youtube Videos Trigger Tingling 'Brain Orgasms' in ASMR Practitioners*. <http://www.nydailynews.com/> (19.06.2025).
- The New London Group 1996, *A Pedagogy of Multiliteracies: Designing Social Futures*, in “Harvard Educational Review” 66 [1], pp. 60-92.
- Thorne S.L. and Hellermann J. 2017, *Mobile Augmented Reality: Hyper Contextualization and Situated Language Usage Events*, in “Proceedings of the XVIII International

- CALL Conference: CALL in Context”, pp. 721-730.
- van Leeuwen T. 1999, *Speech, Music, Sound*, Macmillan, London.
- van Leeuwen T. 2004, *Introducing Social Semiotics*, Routledge, London.
- Wallace M., Woynaroski T. and Stevenson R. 2020, *Multisensory Integration as a Window into Orderly and Disrupted Cognition and Communication*, in “Annual Review of Psychology” 71, 10.1146/annurev-psych-010419-051112.
- Wang Q. 2022, *The Multimodal Construction of Affective Stance in ASMR Videos*, in “Open Access Library Journal” 9, no. e9070.
- Wang Q. 2023, *Digital Intimacy: A Multimodal Discourse Analysis on ASMR Videos*, in “Open Access Library Journal” 10, no. e10538.
- Watts J. 2003, *Politeness*, Cambridge University Press, Cambridge.
- Xin-Yu L. 2022, *The New Transformations and Prospects of Speech Act: Research from the Multimodal Perspective*, in “Journal of Literature and Art Studies” 12 [6], pp. 653-660.
- Zappavigna M. 2019, *Language and Social Media: Enacting Identity through Ambient Affiliation*, in Thompson G., Bowcher W., Fontaine L. and Liang J.Y. (ed.), *The Cambridge Handbook of Systemic Functional Linguistics*, Cambridge University Press, Cambridge, pp. 714-737.
- Zappavigna M. 2021, *Ambient Affiliation in Comments on YouTube Videos Communing around Values about ASMR*, in “Journal of Foreign Languages” 44 [1], pp. 21-40.

HUMOUR AND MODAL INTERPLAY IN TIKTOK'S POV GENRE

AUDREY CLAIRE WILLOUGHBY
UNIVERSITÀ DEGLI STUDI DI MILANO LA STATALE

Abstract - TikTok is widely recognised for its humour-driven performances, yet the specific multimodal strategies through which humour is produced remain underexplored. This study examines humour in the platform's "POV" genre, analysing 16 English-language videos tagged with humour-related hashtags. Drawing on multimodal discourse analysis (Bateman *et al.* 2017; Wildfeuer 2014) and pragmatic humour theory (Attardo 2020; Tsakona 2020), it introduces *cross-modal incongruity* as a distinct form of modal interplay: a moment when one semiotic mode subverts the interpretive frame set by another, prompting the audience to reconcile the mismatch through inference. Six semiotic modes were examined to identify how expectations are established and violated. Cross-modal incongruity was present in every video analysed, most often when textual or verbal framing was overturned through embodied or visual cues. These mismatches acted as prompts for reinterpretation rather than signs of communicative breakdown, drawing on viewers' familiarity with genre conventions, platform performance styles, and shared cultural scripts. The findings suggest that such incongruities are systematically embedded in the multimodal design of POV videos, making the audience's interpretive work integral to the humour and offering a clearer account of how comic effects are orchestrated across audiovisual resources in short-form digital media.

Keywords: multimodal discourse analysis; TikTok; pragmatics; humour; incongruity.

1. Introduction

Short-form video platforms such as TikTok are reshaping how humour is produced and interpreted in everyday digital communication. In highly compressed audiovisual formats, creators draw on semiotic resources such as text overlays, voice, soundtracks, facial expression, gesture, and editing techniques to simulate familiar scenarios and frame social roles. Humour often emerges when these modes clash or shift unexpectedly, producing moments of dissonance that disrupt interpretive coherence and invite incongruous readings (Darvin 2022).

From a pragmatic perspective, humour is not simply a property of a text but a communicative act whose interpretation depends on the interplay between what is said, how it is said, and the shared knowledge and

expectations of participants (Attardo 2020; Tsakona 2020). In this view, incongruity, or violation of expectation, becomes a source of humour through contextualised inferencing, social positioning, and genre competence, which determine whether a multimodal cue is interpreted as signalling the humorous nature of the text.

In TikTok's "point of view" (POV) genre, creators present short fictional scenarios from a first-person or implied perspective, often dramatising everyday dynamics such as parenting, school life, or friendship. These performances rely on rapid sequencing of modes, such as textual overlays, gestures, borrowed media, and audio, where humour often arises from incompatibility in modal alignment that prompt reinterpretation (Bernad-Mechó and Girón-García 2023; Masi 2023). These interferences occur in a temporal sequence, as discussed more below.

While humour in TikTok's POV genre videos often relies on modal shifts that disrupt audience expectations, little research has examined how these violations are constructed and are temporally distributed in short-form audiovisual media. This study addresses that gap by combining multimodal discourse analysis with pragmatic humour theory to investigate cross-modal incongruity as a humour-making strategy, analysing its development over time, across modes, and within narrative positions in relation to contextual expectations.

2. Literature review

2.1. Platform affordances and the POV genre

TikTok's POV genre is shaped by a combination of platform affordances, established performance practices, and platform-specific conventions. In many ways, POV videos build on the logic of static-image memes: while memes condense humour into a single frame, POV videos extend these memetic structures into dynamic, temporally sequenced performances shaped by TikTok's affordances (Shifman 2013; Zulli and Zulli 2020). Traditionally, POV videos were visually framed through the eyes of the character viewers were invited to embody. This literal, first-person usage was present on TikTok from its earliest examples. However, by 2019, the term had broadened to describe videos that invite viewers to engage with a character's perspective or experience without adopting their literal, visual vantage point (Attardo 2025). Both uses have coexisted since that time. In this study, this expanded sense is referred to as the "theatrical" POV, in which scenarios unfold visibly on screen, encouraging relational or emotional identification rather than immersive embodiment.

Platform affordances such as green screen effects, voice dubbing, text

overlays, and audio synchronisation layer meaning: textual overlays can establish roles or narrative frames, while facial expression, camera movement, and sound design set tone and pacing (Divon and Krutrök 2024). As Bucher and Helmond (2018) argue, affordances should not be seen merely as technical features but as action potentials that emerge through interaction between users and the socio-technical environment. On TikTok, these tools enable compact, stylised narratives that draw on cultural scripts to perform irony, exaggeration, or sincerity (Kulaga 2024; Zeng and Abidin 2021; Zulli and Zulli 2020).

The POV format thrives on ambiguity: it gestures toward shared experience while also inviting playful interference. As Trillò (2024) describes, POV videos participate in TikTok's broader "platform vernacular", where repetition, remix, and stylised stance-taking establish community belonging (p. 3). The humour in TikTok's POV genre frequently emerges from viewers' recognition of a familiar interactional script, such as a teacher scolding a class or a couple arguing over something seemingly mundane, and the subtle ways this script is manipulated across modes. Textual overlays frequently establish the expected frame of reference, signalling a character type or scenario, while the subsequent performance introduces dissonance or exaggeration that challenges the viewer's initial assumptions. This approach reflects what Trillò (2024) identifies as a mode of performance in which creators draw on shared cultural scripts to signal affiliation with their audience, while leaving space for irony and playful reinterpretation. Similarly, Divon and Ebbrecht-Hartmann (2022) note that POV videos often organise their narratives and emotional cues through a coordinated use of visual framing, facial expression, and direct camera address, creating a multimodal structure that shapes how viewers interpret and respond to the performance. As Aiello and Parry (2020) explain, these types of shifts are made intelligible through meta-communicative cues that guide interpretation and indicate stance – a process that closely aligns with Gumperz's (1982) notion of contextualization cues, which function similarly to signal speaker stance and frame interpretation as the discourse unfolds.

This use of shared cultural scripts is reinforced by a reliance on memetic structure and repetition, which also contributes to the humour. Shifman (2013), Zulli and Zulli (2020), Vásquez (2019), and Attardo (2023) argue that the viral spread of digital genres is often rooted in repetition-with-variation: creators replicate a known format while introducing subtle shifts in delivery, stance, or outcome. The POV genre benefits from this logic. Users internalise patterns and remix them, signalling affiliation while also inserting playful contradictions. Ironic editing, exaggerated delivery, and visual discontinuities all foster re-interpretation and shared recognition, making humour a product of both performance and platform literacy.

2.2. *Pragmatic incongruity*

Incongruity, typically defined as a clash between what is expected and what is presented, has long been recognised as central to humour (Attardo 1994, 2000; Raskin 1985). From a pragmatic perspective, incongruity functions as a communicative breach: a recognisable mismatch between surface form and inferred meaning (Attardo 2020; Gerrig 1984). In digital contexts, such breaches frequently occur across multiple modes reflecting broader shifts in humour toward fast, interactive, and multimodal formats (Attardo 2023).

This study adopts a pragmatic view of incongruity, where humour arises not only from contrast or contradiction, but also from a viewer's interpretation of intentional misalignment between cues. As Tsakona (2020) points out, humour is always contextually situated as it reflects shared knowledge, genre norms, and social positioning. In the POV genre, creators frequently perform exaggerated versions of familiar roles, using irony and over-performance to stage communicative breaches that may be interpreted as humorous precisely because they deviate from expectations in subtle, stylised ways.

Viewers are expected to infer humorous intent not necessarily from overt jokes or punchlines, but from friction between modes. The idea that humour may stem from unresolved or only partially resolved incongruity that leaves interpretive space open was noted by Attardo and Chabanne (1992) in their analysis of jokes as a text type, and has since been extended to static online formats by Dynel (2016) and Abdel-Raheem (2018), where humour frequently emerges from visual-verbal tension, particularly when image and text combine to set up and subvert expectations. In audiovisual digital communication, such tensions can arise when, for example, a textual overlay suggests a serious scenario that is undercut by an absurd soundtrack or a deadpan reaction. The incongruity lies in the modal misalignment, or in what Holsanova (2014) describes as attentional and interpretive shifts across various modes. These shifts often occur without explicit signalling, relying instead on the viewer's genre literacy and familiarity with platform-specific cues. As Yus (2018) observes, humour in these contexts hinges on the audience's inferential effort to detect relevance among potentially contradictory signs. When viewers recognise that a gaze, gesture, or tonal shift contradicts the expected function of a caption or verbal frame, they resolve this discrepancy through contextual reasoning. The process is dynamic and rarely results in narrative confusion; rather, it invites playful reinterpretation and aligns with the participatory nature of TikTok's platform vernacular. Recognising the breach requires cognitive attention and social competence, lending a sensitivity to genre, irony, and communicative norms.

Humour in POV videos is often built around semiotic violations that subtly interfere with coherence across modes. In this genre, such violations

operate much like the modal misalignments described above, but they are staged within highly compressed, stylised narratives. These moments invite viewers into an active interpretive role, requiring them to identify and reframe unexpected modal shifts as intentional.

2.3. Multimodal approaches to humour

Multimodal discourse analysis (MDA) has received relatively little attention in the study of humour in audiovisual genres, and work on informal, user-generated formats is even more scarce. In particular, few studies have examined the specific role of different modes in creating humorous effects in these contexts. Only recently has MDA been applied to short-form, socially embedded video platforms such as TikTok, where humour circulates in highly participatory and multimodal ways.

Most studies concerned with the modal interplay of online humour have centred on static image–text memes, where juxtaposed modes create contrasts through mismatched tone, framing, or logic (Abdel-Raheem 2018; Dynel 2016; Shifman 2014; Vázquez and Aslan 2021; Yus 2018). While these analyses reveal much about how humour can arise from visual–verbal tension, their focus on fixed, spatial arrangements means the sequence of interpretation is constrained by a single, unchanging visual-textual composition. By contrast, in audiovisual genres, cues from multiple modes unfold, build, and transform as the text progresses. Their sequential organisation actively reshapes expectations over time before they are playfully violated. These genres require analytical approaches that can account for how meaning develops through successive modal shifts, capturing both the temporal organisation of cues and their interaction in producing humour.

Some research has addressed the role of modes in humorous dynamic media: Balirano and Corduas (2008) analyse scripted diasporic television, showing how verbal dialogue, visual framing, and audio cues combine to produce humour rooted in shared cultural knowledge. Bernad-Mechó and Girón-García (2023) examine educational YouTube videos, tracing how humour emerges through the interplay of gesture, the pacing shaped by editing rhythms, and shifts in narrative perspective. While their analysis addresses timing, it centres on the overall structure of a performance rather than on the precise sequencing by which expectations are formed and overturned. Masi (2023) studies how speakers in TED talks use gaze, gesture, prosody, and visual design to create humorous moments through incongruity. Although the study offers valuable insight into multimodal coordination, it gives less attention to how these cues unfold over time to gradually establish and then subvert expectations. These works underscore the importance of orchestrating multiple modes in time yet leave room for closer examination

of the temporal processes through which expectation and violation are constructed.

3. Methodology

3.1. Theoretical approach

This study draws on the works of Bateman, Wildfeuer, and Hiippala (2017) and Wildfeuer (2014), with particular attention to their concepts of modal interplay, interpretive framing, and multimodal coherence. Both frameworks offer structured models for understanding how meaning is constructed across multiple semiotic modes in communicative artefacts.

Bateman *et al.* (2017) describe modal interplay as the ways in which different semiotic modes relate to one another in producing a coherent multimodal discourse. Three primary relations are distinguished in their work: alignment, where modes support the same rhetorical or informational goals; elaboration, where one mode expands, specifies, or adds detail to content presented in another; and contrast, where the content presented in each mode is recognisably different yet integrated into a unified whole. In all cases, coherence is maintained through the integration of modal contributions. Wildfeuer (2014) complements this perspective by emphasising interpretive framing and the viewer's active engagement, showing how filmic discourse guides inference and coherence through shifts in devices such as camera angle, character perspective, or sound design. This focus on the audience's inferential work grounds the present study's extension of both frameworks. Recent work by Zabalbeascoa and Attardo (2023) emphasised that humour should not be understood as confined to linguistic expression, but rather as emerging through semiotic resources more broadly, a shift that underscores the importance of accounting for meaning across modalities.

Building on these perspectives, the current analysis distinguishes modal contrast from *cross-modal incongruity* as possible modal relations that can serve as triggers for humorous interpretation. Whereas modal contrast (as conceptualised by Bateman *et al.* 2017) involves differences between modes that remain structurally coherent and fully integrated into the discourse, offering alternative perspectives or nuances without creating interpretive interference, cross-modal incongruity designates moments in which one or more modes undermine, contradict, or otherwise subvert the interpretive frame established by another. Such instances produce a temporary breakdown in cross-modal coherence, requiring the viewer to resolve the misalignment through pragmatic inference or to interpret it as nonsensical. In this study, only such marked violations of expectations between modes are identified as

cross-modal incongruities and examined as key points in the construction of humour.

This analytical approach is guided by two central research questions:

1. Is cross-modal incongruity central to the humour of TikTok POV videos, and what semiotic strategies are used to achieve it?
2. How are cross-modal incongruities distributed across modes, segments, and narrative positions in humorous POV TikTok videos?

3.2. Data collection

The dataset consists of 16 English-language TikTok videos retrieved via TikTok's API belonging to the "POV" genre, tagged with humour-related hashtags (e.g., #comedy, #funny). From an initial pool of 20 videos, four were excluded: three due to privacy changes and one for having removed the audio.

Metadata such as upload date, like count, hashtags, and captions were recorded but are not analysed in this study. All videos were uploaded between August 2023 and early 2025. Creator demographics were not consistently available and are therefore not addressed.

3.3. Segmentation method

To analyse how expectation and violation unfold to create cross-modal incongruities in TikTok's short-form videos, this study adopts a segmentation method based on interpretive framing (Wildfeuer 2014). Rather than relying on fixed units, such as scenes or time intervals, segmentation follows shifts in the viewer's interpretive orientation prompted by modal cues including gaze, gesture, sound, or camera perspective. Principles from event segmentation theory (Zacks and Swallow 2007) were also integrated, which explains how viewers identify boundaries in unfolding activity based on perceptual salience, such as changes in tone, motion, or visual layout.

A segment is defined here as a bounded stretch in which a stable interpretive frame holds. A new segment is triggered when a shift in one or more modes reorients the viewer's understanding of the communicative act. This framing-based approach is well-suited to TikTok's compressed, stylised format, where coherence depends on rapid multimodal cues and changes in tone or communicative intent.

Each segment is manually annotated in Microsoft Excel, with a new segment identified whenever a semiotic shift is judged to reorganise the viewer's perspective or communicative expectations. This segmentation method enables fine-grained tracking of how expectation is built and then violated across modes, forming the basis for identifying cross-modal

violations and capturing how humour is performed through the temporal interaction of modes.

3.4. Identifying cross-modal incongruity

Each segment was analysed in terms of six semiotic modes: written text, verbal spoken language, audio (such as music or sound effects), visual design (including background, editing choices, setting, and props), gesture, and facial expression (including gaze). For each mode, meaning was paraphrased, and segments were flagged for incongruity when they showed an interference in interpretive alignment either within a single mode or between modes. In these cases, the analysis identified which modes established the initial expectation (Meaning 1) and which violated said expectation (Meaning 2). Cross-modal incongruity was annotated when the mode or mode combination establishing the initial expectation differed from the mode or mode combination producing the violation, for example, when a joyful textual overlay was followed by a flat, ironic delivery, or when a facial expression contradicted the tone of spoken language. Incongruity could also occur within a single mode, such as a textual overlay that itself contained a tonal or logical mismatch. The aim was not to classify types of humour, but to identify recurring patterns in how audiovisual resources violate expectation.

3.5. Examples: Segmentation and cross-modal incongruity

3.1.5. Example one: V_001



Note. S: – POV student speech, T: – teacher speech

Figure 1
Video V_001.

Figure 1 demonstrates how a video from the data set was segmented. Video V_001 provides a representative case of how segmentation and violation of expectations are used to identify cross-modal incongruity. The video features a 23-second performance titled “POV: That one teacher during an exam”, in which a single speaker enacts a teacher figure with exaggerated gestures and classroom mannerisms. Across ten segments, the video gradually builds an interpretive frame of a strict, mildly passive-aggressive authority figure through the interplay of textual overlay, gaze, gesture, and speech.

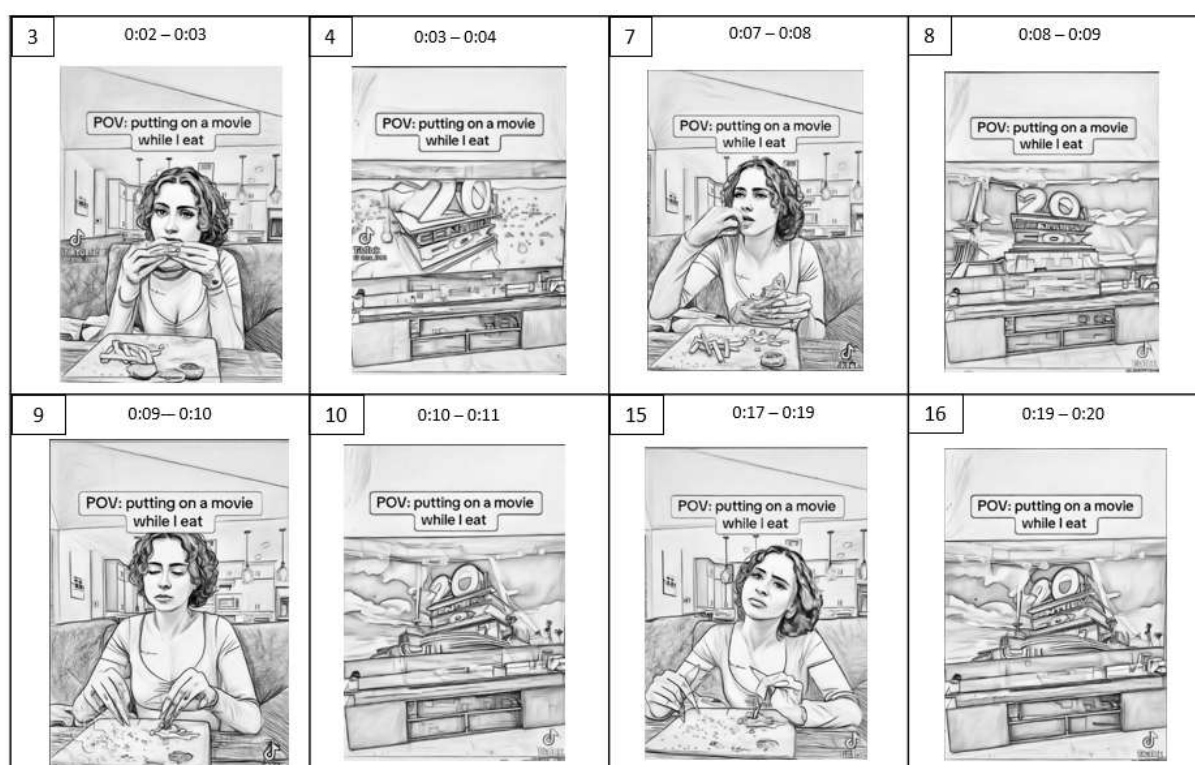
The first three segments are non-verbal: The video opens to the teacher scanning the classroom in silence with pursed lips as she crosses her arms, folding her sweater across her chest. At second 0:03 she tilts her head and makes eye contact with the camera/POV student, indicating a new segment. Around second 0:06 her attention turns to the POV student's work, as she also raises her eyebrows and upturns her mouth, indicating judgment and disapproval at what she sees, signalling the beginning of the third segment. The meaning-making across modes within these segments works in close alignment with the initial POV textual overlay by elaborating on one another to reinforce a consistent scenario, role, and tone.

Segments 4 through 7 each mark subtle reorientations. Segment 4 introduces the student's discomfort when he says, “Um, hello...”, drawing attention to the teacher's continued silence. Segment 5 begins with the teacher's first spoken performance in the video, as she pulls away from the POV student and addresses the entire class, instructing everyone to double-check their work after viewing the POV student's exam. Segments 6 and 7 elaborate on this frame as she continues in a lecturing tone, with her index finger raised, to mildly scold about the performance of the test. These changes in speech act, gaze direction, and facial expression justify new segment boundaries, since perceptual shifts of this kind cue the viewer to register a boundary in unfolding activity (Zacks and Swallow 2007) and prompt an update in how the ongoing event is mentally represented. At the same time, they function as interpretive framing devices in Wildfeuer's (2014) sense, guiding the viewer's understanding of communicative intent.

A single, clear incongruity occurs across segments 7 and 8. In Segment 7, the teacher delivers the line: “I'm not gonna name names” with her hands raised, palms outward, as clearly seen in the corresponding visual frame in Figure 1. These verbal and gestural cues set an expectation of discretion, aligning with the role of a strict but fair authority figure. However, in Segment 8, she makes deliberate and prolonged eye contact with the camera (positioned as the POV student) during an awkward pause, hands still raised, thereby undermining the very expectation she just constructed by implying who the student-in-question actually is. This breach is not accidental or narratively repaired, but rather staged as an ironic, stylised contradiction between modes.

This is the only moment in the video that qualifies as a cross-modal incongruity. The humour arises from the viewer recognising the mismatch between verbal intent and gaze delivery. Rather than being explicitly resolved narratively, this incongruity remains open-ended, inviting pragmatic reinterpretation. In Segments 9 and 10, the teacher resumes her authoritative tone, returning to addressing the entire class without acknowledging the earlier interference. Although unresolved within the video's explicit narrative, viewers readily interpret this violation of expectation as humorous, recognising it both as a deliberate, exaggerated performance typical of the POV genre and as a familiar behavioural pattern. The humour, in other words, is not only triggered by the incongruity itself but also by the recognition of a familiar situation now being made visible and exaggerated for comic effect. This example thus demonstrates how segmentation based on modal shifts and interpretive reorientation can capture humour that emerges from the interplay of carefully timed intermodal contradictions and culturally shared recognition.

3.2.5. Example two: V_007



Note. Not all 16 segments of V_007 are present in Figure 2 due to space constraints; only 8 were selected to demonstrate the use of specific modes to convey the passage of time, as discussed below. The segment number out of the 16 total is listed in the upper left corner of each visual frame.

Figure 2
Video V_007.

Video V_007 is a 20-second montage of 16 segments with no spoken dialogue, alternating between shots of a young woman eating at a coffee table in her living room and shots of the modern 20th Century Fox film intro playing on her television. The actual length of the Fox intro is approximately 20 seconds, the same length as this TikTok. Yet through the montage structure, the viewer perceives it as lasting far longer: the woman's food visibly diminishes at a normal eating pace, creating the impression that more time has passed than the actual 20-second duration of the intro. The music of the intro plays continuously and at normal speed across cuts, meaning it is heard even when the camera cuts to the young woman eating, and the visual movements of the intro continue to unfold whenever the camera returns to the TV.

The video opens with the textual overlay "POV: putting on a movie while I eat", framing the scenario and establishing *meaning 1*: the expected sequence of eating a meal while a film plays. In segment 1, she bites into a burger, fries and chicken nuggets on the plate beside her. Segment 2 cuts to the Fox intro, its animation and music progressing normally as mentioned above. Segment 3 returns to her, taking a large bite; the burger is visibly smaller than it was in segment 1. Segment 4 cuts back to the Fox intro, still unfolding. This alternation continues: segment 5 shows her eating a chicken nugget; segment 6 again cuts back to the intro, progressing; segment 7 shows the burger now halved, the nuggets gone, and her focus shifting to the fries. Segment 8 again shows the intro's movements and music continuing. As this sequence develops, the embodied mode of eating and the visual cue of disappearing food begin to suggest that more time is passing between cuts than the real-time unfolding of the intro would allow, elongating its perceived duration.

By segment 9, most fries, and therefore almost the entire meal, are gone, and she glances down at her plate, followed by segment 10, in which the intro is still in motion. This is the point where *meaning 2* emerges: the expectation of enjoying a meal during a movie is violated because the meal ends before the film even begins. The incongruity could, in theory, also be read as her eating unusually fast. Still, because her on-screen eating is shown at a normal pace, inference favours the first interpretation - that the intro in the narrative reality of the TikTok is absurdly long.

Segments 11 and 12 reinforce the violation: the final bites of fries are eaten, the plate is empty, yet the intro continues unfolding on-screen. In segment 13, she wipes crumbs from her hands, an embodied cue of completion, followed by segment 14, yet another view of the still-progressing intro. Segment 15 delivers the reaction: she rests her arms on the table, tilts her head toward the TV, furrows her brow, and blinks in exaggerated disbelief. Segment 16 closes with one final cut to the intro as it continues still.

Embodied cues make the violation visible: the diminishing food portions, the hand movements signalling eating and then completion, and her closing reaction all contrast with the uninterrupted progression of the intro. The music's steady pace and the fluid motion of the intro confirm that no time compression or acceleration is happening within the intro itself. The montage thus overlays two timelines whose durations become incompatible in the viewer's mind: the completion of a meal versus the start of a film.

The humour here depends on the alignment and later divergence of modes. Textual framing, embodied action, and visual details of the disappearing meal set and sustain the initial expectation, while the visuals and audio of the 20th Century Fox intro establish the competing temporal frame. The violation lies in their sustained mismatch, a relatable recognition of how long movie intros can feel, reframed here through an exaggerated scenario.

As Attardo (2020) argues, the processing of incongruity depends on inferential mechanisms tied to what the viewer already knows about a character, context, and genre. In both examples, viewers recognise the contradictions not as communicative failures, but as deliberate cues for humour, interpreting them through familiarity with the POV format's exaggerated roles and performance conventions. Tsakona (2020) similarly emphasises that humour is shaped by genre conventions and social positioning, both of which inform how audiences read the teacher's performative breach in V_001 and the temporal absurdity of V_007. In each case, the incongruity remains unresolved within the narrative: in V_001, the teacher's direct gaze at the "unnamed" student is left hanging, while in V_007, the never-ending movie intro continues without closure. These deliberate non-resolutions work as signals to the audience that the contradiction itself is the point of amusement, encouraging them to actively interpret and engage with the performance.

4. Results and discussion

4.1. *Cross-modal incongruity in the dataset*

The analysis of 16 TikTok POV videos reveals that multimodal incongruity is a fundamental mechanism in how humour is constructed within this dataset. In each video, at least one instance was observed in which two or more semiotic modes collaboratively established an expectation, which was then violated another mode, creating a moment of interpretive interference. These instances were coded as occurrences of cross-modal incongruities and form the core of the analysis presented here.

4.2. Frequency and distribution of incongruity

A total of 28 instances of incongruity were identified across the 16-video dataset (see Table 1), 26 of which cross-modal incongruities, and 2 monomodal incongruities.

Video	Video length	# of segments	# of incongruity	Segment in which incongruity/ies occur(s)
V 001	00:23	9	1	7,8
V 002	00:15	6	3	1, 2, 6
V 003	00:10	2	2	1*, 1
V 004	00:07	4	1	1
V 005	00:10	2	1	2
V 006	00:14	6	1	3
V 007	00:20	16	2	9,10
V 008	00:19	9	3	4, 8, 9
V 010	00:07	4	1	3
V 011	00:11	7	2	2, 6
V 012	00:06	5	2	1*, 1
V 013	00:06	5	2	2, 4
V 014	00:07	2	1	1
V 016	00:14	4	1	1
V 017	00:06	2	1	1
O 019	00:14	10	4	5, 6, 8, 10

Note. Only 2 of the 28 instances of incongruity were monomodal, indicated with an astrisk (*) in Table 1.

Table 1
Incongruity occurrences and segmentation across 16 videos in dataset.

All 16 videos in the dataset contained at least one instance of incongruity, distributed as follows:

- 8 videos featured one instance of cross-modal incongruity
- 3 videos featured two instances of cross-modal incongruity
- 2 videos featured three instances of cross-modal incongruity
- 1 video featured four instances of cross-modal incongruity
- 2 videos featured two instances of incongruity: one cross-modal and one monomodal textual incongruity

This distribution suggests that while many creators build toward a single key moment of modal interference, others layer multiple incongruities throughout the video, constructing more complex humorous arcs. Attardo's (2001) distinction between jablines and punchlines helps account for this variation: some incongruities operate like jablines, dispersed across segments as secondary humorous turns, while others resemble punchlines, concentrated in the final position as climactic resolutions. In this sense, the segmental positioning of incongruities becomes central to understanding how humorous effects are organised within POV clips.

The consistent presence of at least one instance of multimodal incongruity in every video further indicates that intermodal expectation

violation is a defining feature of humorous performance in the POV genre. Monomodal incongruities, while rarer, still contribute to the overall logic of humour in this genre by destabilising the established frame or subverting expectations within the same communicative channel.

4.3. Segmental positioning of incongruity

Occurrences of multimodal incongruity appeared at different narrative points but most frequently in the middle or later segments, following the establishment of a stable interpretive frame. For example:

- In V_001, cross-modal incongruity emerges across segments 7 and 8, where bodily and tonal shifts undermine the verbal premise
- V_008 presents three instances of cross-modal incongruity across segments 4, 8, and 9, suggesting a pacing strategy of escalating interference
- In V_011, a cross-modal incongruity occurs already in segment 2 out of 7, demonstrating that humour can also arise early in the video if enough contextual framing is established in the opening moments

In contrast, seven videos feature a cross-modal incongruity immediately in segment 1. In six of these, creators recontextualise intertextual media clips (remix), such as scenes from television shows, movies, or other viral videos, as the audiovisual base for their POV scenarios. These clips are paired with new textual overlays that assign an interpretive frame often at odds with the original content. The result is an immediate incongruity between the expectation created by the textual overlay and the meaning created by the audiovisual material.

For instance, in V_004, the textual overlay reads “POV: me trying to study”, setting up an expectation of studiousness and focused effort. In contrast, the accompanying video clip from *Ice Age 2* shows Sid the Sloth flailing on the ground and making guttural, nonsensical noises. The textual mode and the audiovisual clip belong to entirely different semantic domains: the text evokes discipline and concentration, while the moving image and sound convey chaos, absurdity, and lack of control. This stark mismatch produces an initial interference in interpretive alignment. Viewers resolve the interference by reframing the chaotic clip as a metaphor for mental disorganisation or the difficulty of maintaining focus – a familiar scenario to anyone who has ever tried to study – so that the incongruity ultimately functions as a source of humour.

These immediate contrasts rely less on narrative buildup and more on the viewer’s ability to resolve cross-modal incongruity in the moment, interpreting the misalignment through inference and cultural reference. In such cases, the humorous effect emerges not from progression, but from the

interplay between media recognition and textual framing.

4.4. Modal roles in expectation and violation

Each instance of incongruity was analysed by identifying which mode(s) contributed to the initial expectation (Meaning 1) and which violated it (Meaning 2). A breakdown of these interactions per video is shown in Table 2.

Video	Meaning 1 (Expectation)	Meaning 2 (Violation)
V 001	verbal	Expression
V 002	text	Visual
V 002	text	Verbal
V 002	text	verbal, visual
V 003	text	Text
V 003	text	Audio
V 004	text	audio, visual
V 005	gesture	audio, gesture
V 006	verbal	Expression
V 007	gesture, text, visual	audio, visual
V 007	gesture, text, visual	audio, visual
V 008	gesture	Expression
V 008	gesture	Expression
V 008	gesture, audio	expression, gesture
V 010	verbal	expression, gesture
V 011	verbal	Expression
V 011	verbal	Expression
V 012	text	Text
V 012	text	expression, visual
V 013	gesture, text	Audio
V 013	text	Gesture
V 014	text	Visual
V 016	text	verbal, visual
V 017	text	verbal, visual
V 019	visual	expression, gesture,
V 019	visual	expression, gesture
V 019	visual	audio, expression, gesture
V 019	audio, text, visual	audio, expression, gesture

Table 2
Expectation and Violation Modes per Video.

The distribution of modal roles across the 28 instances of incongruity reveals consistent patterns in how expectation and violation are constructed, outlined more clearly in Tables 3 and 4 below.

Mode	Occurrences
text	16
gesture	7
visual	6
verbal	5
audio	2
expression	0

Note. Occurrences across the entire data set.

Table 3
Meaning 1 – Expectations.

Mode	Occurrences
expression	13
visual	9
gesture	8
audio	8
verbal	4
text	2

Note. Occurrences across entire data set.

Table 4
Meaning 2 – Violations.

Textual overlays led as the primary framing mode to set up expectations that are subsequently violated (16 instances) as seen in Table 3. Gesture (7) and visual (6) modes were the second most common expectation-setting modes, often adding specificity to the scenario or emphasising the tone implied by the textual overlay. Verbal speech followed with 5 occurrences, typically reinforcing or elaborating the role or premise established in the overlay. Audio was comparatively rare in establishing expectations (2 occurrences), but when present, it often invoked recognisable media references or environmental cues that anchored the setting. These figures suggest that linguistic cues, particularly the POV textual overlay, operate alongside embodied and visual signals to construct the viewer's interpretive frame, signalling character identity, social context, or emotional tone, and preparing the viewer to read the scene in a particular way.

In contrast, Table 4 demonstrates that violations were most commonly enacted through expressive and embodied modes. Facial expression contributed to Meaning 2 in 13 instances, visual cues in 9 instances, and gesture and audio in 8 instances each. Verbal interferences occurred in 4 cases, while text appeared only twice, as part of a violation, and in both instances, it was the only monomodal incongruity in the dataset. The cross-modal incongruities in this dataset frequently took the form of embodied or visual interference to an established linguistic premise, creating a recognisable gap between what is verbally or textually implied and what is

shown or performed.

This recurring structure, in which linguistic framing is destabilised by visual or embodied modes, sets the stage for a broader theoretical reflection. It appears to support, at least in the context of POV humour, Lotman's (1975) proposal that language functions as a "primary modelling system" through which other semiotic systems derive coherence. In these examples, linguistic elements often establish the initial meaning configuration, while subsequent modalities introduce shifts that reframe or violate that configuration. By comparison, other TikTok genres, such as dance trends, visual memes, or reaction videos, may open with non-linguistic modes like movement, sound, or facial expression to set the premise. This variation in which mode initiates meaning shows that modal salience is shaped by genre, supporting multimodal discourse theories (Bateman *et al.* 2017; Jewitt 2014; Kress and van Leeuwen 2001) that view modes as operating through shared semiotic principles rather than presupposing inherent communicative dominance of one mode over others.

In the case of POV humour, linguistic and visual scaffolding is often deliberately set up so that it can be undermined. This aligns with Tsakona's (2020) view of humour as a pragmatic and context-dependent phenomenon shaped by shared genre expectations. Creators in this dataset draw on familiar social archetypes, such as teachers, parents, or work-related situations, which are rendered legible through text, gesture, or visual detail, and then exaggerated or subverted through facial expression, tone, or visual reversal. The segmentation method used in this study was instrumental not only in isolating these moments of cross-incongruity but also in capturing how they unfold over time and across different modes. This offers a practical model for identifying humour that is performed and modally layered, rather than explicitly stated.

5. Conclusion

This study shows that cross-modal incongruity is a deliberate and consistent humour-making strategy in TikTok's POV genre. In every video analysed, at least one expectation was established in one mode, then later violated by another, most often when textual or verbal framing was overturned through facial expressions, gestures, or visual details. These moments invited viewers to reinterpret the scene in light of the new cue, recognising the misalignment as intentional and drawing on genre knowledge, platform norms, and shared cultural scripts to resolve it as humorous. Monomodal incongruities were rare in this dataset, suggesting that, in these examples, humour was typically achieved through the interplay of modes rather than through a single mode acting alone. While the limited sample prevents broader generalisations, this

tendency highlights how even within very short clips, creators often mobilise multiple semiotic resources in tandem, making the multimodal orchestration of incongruity a particularly salient feature of the humorous effect.

The segmentation approach was central to revealing how these incongruities function. By locating each violation within the sequence of frame-building and frame-breaking moments, it became possible to see whether humour was constructed through gradual layering or introduced immediately, as in remix formats. This temporal mapping also showed that leaving an incongruity unresolved often positioned the viewer's act of reinterpretation as part of the joke. Such insights would not emerge as clearly from analysing moments of tension in isolation, since the timing and relation to the preceding frame are key to how the humour is perceived.

Overall, the findings highlight a genre-specific strategy in which linguistic framing establishes an initial reading that is later playfully violated through embodied or visual modes. Understanding this modal interplay offers a clearer view of how humour in participatory digital media depends on both the precise coordination of modes and the inferential engagement of audiences. Future work could investigate whether similar strategies operate in other TikTok genres or how audiences from different cultural contexts recognise and resolve these incongruities, further clarifying the social and cognitive processes that make them effective.

Bionote: Audrey Willoughby is a doctoral candidate in Linguistics at the University of Milan, in the Department of Languages, Literatures, Cultures and Mediations. She earned her Master's Degree in Applied Linguistics from Texas A&M University-Commerce, where she wrote her thesis *Humor Markers in Computer-Mediated Communication*, and completed a Graduate Certificate in TESOL. Her doctoral research examines how humour is realised semiotically in short-form video platforms, drawing on multimodal discourse analysis to examine meaning-making in digital performance media.

Author's address: audrey.willoughby@unimi.it

References

- Abdel-Raheem A. 2018, *Multimodal Humour: Integrating Blending Model, Relevance Theory, and Incongruity Theory*, in "Multimodal Communication" 7 [1], pp. 1-25.
- Aiello G. and Parry K. 2020, *Visual Communication: Understanding Images in Media Culture*, SAGE Publications, London.
- Attardo S. 1994, *Linguistic Theories of Humor*, Mouton de Gruyter, Berlin.
- Attardo S. and Chabanne J.C. 1992, *Jokes as a Text Type*, in "Humor" 5 [1/2], pp. 165-176.
- Attardo S. 2020, *The Linguistics of Humor: An Introduction*, Oxford University Press, Oxford.
- Attardo S. 2001, *Humorous Texts: A Semantic and Pragmatic Analysis*, Mouton de Gruyter, Berlin.
- Attardo S. 2023, *Humor 2.0: How the Internet Changed Humor*, Anthem Press, London.
- Attardo S. 2025, *The Shifting Semantics of Point of View (POV)*, in "Salvatore Attardo Blog", <https://salvatoreattardo.substack.com/p/the-shifting-semantics-of-point-of> (19.9.2025).
- Balirano G. and Corduas M. 2008, *Detecting Semiotically-Expressed Humor in Diasporic TV Productions*, in "Humor" 21 [3], pp. 227-251.
- Bateman J.A., Wildfeuer J. and Hiippala T. 2017, *Multimodality: Foundations, Research and Analysis – A Problem-Oriented Introduction*, De Gruyter Mouton, Berlin.
- Bernad-Mechó E. and Girón-García C. 2023, *A Multimodal Analysis of Humour as an Engagement Strategy in YouTube Research Dissemination Videos*, in "The European Journal of Humour Research" 11 [1], pp. 46-66.
- Bucher T. and Helmond A. 2018, *The Affordances of Social Media Platforms*, in Burgess J., Marwick A. and Poell T. (ed.), *The SAGE Handbook of Social Media*, SAGE Publications, London, pp. 233-253.
- Gerrig R.J. 1984, *On the Pretense Theory of Irony*, in "Journal of Experimental Psychology: General" 113 [1], pp. 121-126.
- Darvin R. 2022, *Design, Resistance and the Performance of Identity on TikTok*, in "Discourse, Context & Media" 46 [100591], pp. 1-11.
- Divon T. and Eriksson Krutrök M. 2024, *Playful Trauma: TikTok Creators and the Use of the Platformed Body in Times of War*, in "Social Media + Society" 10 [3], pp. 1-15.
- Divon T. and Ebbrecht-Hartmann T. 2022, *Performing Death and Trauma? Participatory Mem(e)ory and the Holocaust in TikTok #POVChallenges*, paper presented at AoIR 2022: The 23rd Annual Conference of the Association of Internet Researchers, Dublin, Ireland, November 2022.
- Dynel M. 2016, *"I has seen image macros!": Advice Animal Memes as Visual-Verbal Jokes*, in "International Journal of Communication" 10, pp. 660-688.
- Gumperz J. 1982, *Discourse Strategies*, Cambridge University Press, Cambridge.
- Holsanova J. 2014, *Reception of Multimodality: Applying Eye Tracking Methodology in Multimodal Research*, in Jewitt C. (ed.), *The Routledge Handbook of Multimodal Analysis*, 2nd ed., Routledge, London, pp. 285-296.
- Jewitt C. 2014, *An Introduction to Multimodality*, in Jewitt C. (ed.), *The Routledge Handbook of Multimodal Analysis*, 2nd ed., Routledge, London, pp. 15-30.
- Kress G. and van Leeuwen T. 2001, *Reading Images: The Grammar of Visual Design*, Routledge, London.

- Kuřaga W. 2024, *Revolutionizing Visual Communication and Digital Creative Engagement: The Game-Changing Impact of TikTok*, in “Przegląd Socjologii Jakościowej” 20 [3], pp. 212-235.
- Lotman Y.M. 1975, *On the Metalanguage of a Typological Description of Culture*, in “Semiotica” 15 [2], pp. 97-123.
- Masi S. 2023, *Humour in TED Talks: A Multimodal Account*, in “ESP Today” 11 [2], pp. 328-348.
- Raskin V. 1985, *Semantic Mechanisms of Humor*, Reidel, Dordrecht.
- Shifman L. 2013, *Memes in Digital Culture*, MIT Press, Cambridge (MA).
- Shifman L. 2014, *The Cultural Logic of Photo-Based Meme Genres*, in “Journal of Visual Culture” 13 [3], pp. 340-358.
- Trillò T., 2024, “PoV: You are Reading an Academic Article”. *The Memetic Performance of Affiliation in TikTok’s Platform Vernacular*, in “New Media & Society” 0 [0], pp. 1-22.
- Tsakona V. 2020, *Recontextualizing Humor: Rethinking the Analysis and Teaching of Humor*, De Gruyter Mouton, Berlin.
- Vásquez C. 2019, *Language, Creativity and Humour Online*, Routledge, London.
- Vásquez C. and Aslan E. 2021, “Cats be outside, how about me?”: *The Linguistic Construction of Stance in Digitally Mediated Humorous Narratives*, in “Journal of Pragmatics” 171, pp. 240-253.
- Wildfeuer J. 2014, *Film Discourse Interpretation: Towards a New Paradigm for Multimodal Film Analysis*, Routledge, London.
- Yus F. 2018, *Multimodality in Memes: A Cyberpragmatic Approach*, in Bou Franch P. and Garcés Conejos Blitvich P. (ed.), *Analyzing Digital Discourse: New Insights and Future Directions*, Palgrave Macmillan, London, pp. 105-131.
- Zabalbeascoa P. and Attardo S. 2023, *Humour Translation Theories and Strategies*, in L. Kostopoulou and V. Misiou (ed.), *Transmedial Perspectives on Humour and Translation: From Page to Screen to Stage*, Routledge, London/New York, pp. 13-32.
- Zacks J.M. and Swallow K.M. 2007, *Event Segmentation*, in “Current Directions in Psychological Science” 16 [2], pp. 80-84.
- Zeng J. and Abidin C. 2021, “#OkBoomer, Time to Meet the Zoomers”: *Studying the Memefication of Intergenerational Politics on TikTok*, in “Information, Communication & Society” 24, pp. 2459-2481.
- Zulli D. and Zulli D.J. 2020, *Extending the Internet Meme: Conceptualizing Technological Mimesis and Imitation Publics on TikTok*, in “New Media & Society” 24 [3], pp. 614-636.

AI-POWERED MULTIMODAL FEEDBACK IN EFL WRITING

Pragmatic Functions and Classroom Implications

ANTONELLA GIACOSA
UNIVERSITY OF TURIN

Abstract – Feedback is a pragmatically complex aspect of English as a Foreign Language (EFL) pedagogy, as it includes various speech acts such as evaluating, giving advice, correcting errors, and balancing praise with criticism (Liu *et al.* 2021), which can significantly influence student motivation and progress (Kerr 2020). This study adopts an exploratory mixed-methods design that integrates three datasets to investigate multimodal feedback in EFL writing classes through the dual lenses of teachers' pragmatic awareness and students' perceptions. The research focuses on an experimental activity involving fifty secondary school students' (B1–C1) use of Grammarly®, an AI-powered writing assistant. This tool offers automated written corrective feedback (AWCF) by providing scores, colour-coded suggestions, and visual cues related to grammar, spelling, style, and tone. By simultaneously evaluating, encouraging and correcting students, AWCF can help teachers blur the traditional lines between formative and summative feedback (Kerr 2020). The three datasets consist of: the classification of AWCF based on pragmatic functions (explanation, evaluation, suggestion, encouragement) (Liu *et al.* 2021) and realisation modes (linguistic, visual, interactive) (Herring 2019; Kress and van Leeuwen 2020); anonymous questionnaires assessing students' perceptions before and after the activity; and a self-evaluation of EFL teachers' awareness of multimodal feedback through an anonymous questionnaire. In addition to confirming its effectiveness and motivational potential, the research suggests that AWCF may serve as a multimodal, pragmatic partner for teachers: by offering more consistent, transparent, and interactive feedbacks through the semiotic modes of text, visuals, and numerical scoring. The findings suggest that such tools may help instructors expand Classroom Interactional Competence (CIC) through integrative awareness and reflective adaptability to students' needs.

Keywords: feedback; pragmatic awareness; multimodality; classroom interaction competence (CIC); EFL.

1. Introduction

Feedback is a crucial element in English as a Foreign Language (EFL) classes, significantly influencing student motivation and progress (Kerr 2020). By providing learners with information on their skills and proficiency, it helps them detect their progress, become aware of their strengths and weaknesses and identify strategies for improvement (Nava and Pedrazzini

2018). Generally, instructors provide summative feedback during evaluations, while also striving for formative outcomes aimed at enhancing students' skills and proficiency (Ajjawi *et al.* 2021). Nonetheless, students often perceive feedback as judgmental, interpreting it not only as an assessment of performance, but also as a reflection on themselves personally (Ismail *et al.* 2022). As a result, feedback may be viewed as discouraging, thus failing to support learner development and adversely affecting their motivation and self-efficacy (Carless and Winstone 2020).

Providing feedback is a complex task that requires effective pragmatic skills, as it involves various speech acts such as evaluating, giving advice, correcting errors, and balancing praise with criticism (Liu *et al.* 2021). Teachers need to consider form, intent, and tone in addition to learner language proficiency, cultural backgrounds, and emotional states (Darong and Guna 2023; El-Dakhs *et al.* 2019). This pragmatic competence draws on professional awareness and plays a crucial role in teachers' Classroom Interactional Competence (CIC). This term refers to the strategic use of interaction to foster understanding and engagement and enhance learning (Walsh 2014). The ways teachers align form – spoken, written, or multimodal – intent, and tone in their feedback directly shape student engagement, comprehension, motivation, and classroom climate (Ishihara 2011; Ekin and Damar 2013).

Whereas classroom discourse once relied on traditional, monomodal approaches (such as written or oral feedback), there is now a shift toward more dynamic and integrated practices that better align with 21st-century learning needs (Steiss *et al.* 2024). In addition to conventional means of providing feedback, educators can rely on technologically enhanced feedback, including digital and video-based feedback and multimodal annotations (Cunningham 2019). A growing body of research has examined AI as an automated generator of writing feedback, often referred to as Automated Written Feedback (AWF) (Huawei and Aryadoust 2024). Recently, AI-powered tools such as Grammarly® have been investigated as generators of automated written corrective feedback (AWCF), which provides advanced feedback on written text, surpassing traditional spelling and grammar checkers (Ranalli and Yamashita 2022). There is no clear consensus among researchers regarding the positive impact of AWCF on student learning outcomes. On the one hand, integrating AWCF into EFL writing instruction does not necessarily result in measurable improvements in student writing quality (Fan 2023). On the other hand, AWCF has proved as a significant element in teaching practices aimed at developing innovative teaching methods to improve student writing abilities (Youn *et al.* 2025). By providing multimodal feedback, AI-powered tools can help shift from a conventional result-oriented feedback model to more student-centred practices, allowing teachers to focus on higher-level writing concerns like

content and organisation (Bayat *et al.* 2020; Farnia *et al.* 2025). Although AI-generated feedback could represent a viable alternative or supplement to human-provided input, a blended approach of human and AI-generated feedback is recommended as the most effective approach, since it leverages the unique strengths of each (Escalante *et al.* 2023). However, despite the need for sharper digital interactional skills (Moorhouse *et al.* 2021), AI-generated multimodal feedback remains underrepresented in research on teacher feedback literacy (Fan 2023; Jiang *et al.* 2024)

To contribute to the body of knowledge on AI for educational purposes, this mixed methods research focuses on AWCF as a means of enhancing teacher pragmatic skills in EFL writing classes. It investigates AI-generated feedback on student writing from multimodal perspective and considers teacher pragmatic awareness and student perceptions. As outlined in the methodology section, below, the study focuses on an experimental activity involving two intact secondary school classes (N=50) who used Grammarly¹ as an AI-powered AWCF feedback tool. This commercial AWCF service was chosen because it is freely available, globally accessible, and widely used, which makes the findings applicable to a broad user base. The analysis explores how the multimodal features of AWCF can help teachers blur the traditional lines between formative and summative feedback. The findings discuss how AI-powered multimodal feedback can broaden teacher pragmatic competence while sustaining student motivation and improving writing skills.

2. Methodology

2.1. Logic of inquiry and methods of data collection

This exploratory study references previous research based on Austin's and Searle's speech act theory, which views teacher feedback as a series of illocutionary acts intended to correct, guide, praise, or criticise student work with the goal of enhancing comprehension and proficiency (Alimova *et al.* 2025; Hyland and Hyland 2001). Against the backdrop of the recent technological turn in education, the study investigates how AI tools like Grammarly can incorporate teacher pragmatic competence. Specifically, it examines:

¹ This study employed Grammarly version 14.1241.0, released in mid-June 2025, rather than the most recent version 14.1244.0, which became available on July 10, 2025, after data collection had been completed. This is not considered a limitation of the study, as the functions analysed align with those commonly referenced in the existing literature.

- The multimodal and pragmatic features of Grammarly's AWCF (Research question 1)
- Students' perceptions of how AWCF affects their writing (Research Question 2)
- Teachers' reflections on their feedback practices, including pragmatic competence and familiarity with AI writing tools (Research Question 3)

As displayed in Table 1, this study adopts a convergent mixed methods design, concurrently collecting quantitative and qualitative data from three sources (Creswell and Creswell 2018). The data examined in this study include: 1) the classification of Grammarly's feedback based on pragmatic functions (evaluation, explanation, encouragement) (Liu *et al.* 2021) and realisation modes (linguistic, visual, interactive) (Herring 2019; Kress and van Leeuwen 2020); 2) anonymous questionnaires assessing students' perceptions of the activity's impact; and 3) a self-evaluation of teachers' awareness of multimodal feedback, collected through anonymous questionnaires administered to EFL teachers. This procedure generated both qualitative and quantitative data, yielding three complementary datasets, which were analysed separately and then compared to provide a deeper understanding of how AI-powered tools may enhance teachers' pragmatic effectiveness and students' writing skills.

Data for this study were collected from two intact secondary school writing classes taught by the same instructor, comprising a total of 50 students (60% reported a B1 level and 40% a C1 level of English proficiency). In Step 1, prior to the writing class, students wrote a 100-150-word text on a given prompt as a home assignment and submitted it via Google Forms. In Step 2, the teacher-researcher randomly selected 20 texts and processed them in Grammarly to observe the type of feedback provided. AWCF was obtained through the free version of Grammarly, with the interface language set to English and default settings applied for domain, tone, and writing goals. The "Insights" panel, displaying overall performance statistics and suggestions, was available to the user. AWCF relating to these texts was then classified for analysis, combining a multimodal and pragmatic perspective (Dataset 1, see 3.2). Step 3 refers to the writing class itself, where the participants used the AWCF provided by the AI-powered tool to edit the texts that they had previously written, under the teacher-researcher's supervision. At the end of the lesson (Step 4), students were given a link to an anonymous questionnaire designed to help them reflect on the type of feedback they received from the AI-powered tool. The questionnaire included 26 questions (a mix of open- and closed-ended questions) investigating students' needs regarding feedback and their experience with AWCF (Dataset 2). Finally, to gain a broader perspective, ten secondary school EFL teachers were invited to complete an anonymous questionnaire on their feedback

practices and their experiences with, and opinions of, AWCF (Dataset 3). All participants had over ten years of EFL teaching experience but displayed differing attitudes toward AWCF. Six reported having no experience with AWCF; three used it primarily to edit their own writing; and only one had introduced it to students. Despite the limited number of teacher participants, this group is relevant as it reflects the level of AI familiarity of secondary school teachers and lack of integration of AI into pedagogy, issues also reported in recent studies on teachers' readiness and attitudes toward AI integration in education (Moorhouse 2024; van den Berg 2025).

Step	Description	Dataset and data type	Research questions
1	Home assignment Students write and submit a 100–150-word text		
2	20 texts (10 from each group) are randomly selected and run through Grammarly	Dataset 1 QUAL+ QUANT for multimodal and pragmatic analysis of 20 students' texts	RQ1 What multimodal and pragmatic features of AWCF may students receive?
3	Writing class		
4	Perceptions of feedback tone, helpfulness, motivation	Dataset 2 QUAL+ QUANT for thematic analysis of student questionnaire responses	RQ2 How do students perceive AWCF?
	Views on pedagogical value, alignment with classroom feedback	Dataset 3 QUAL+ QUANT for thematic analysis of teacher questionnaire responses	RQ3 How do teachers perceive their skills regarding feedback and the use of AI?

Table 1
Logic of inquiry.

2.2. Methods of data analysis

2.2.1. Dataset 1: A pragmatic overview of AWCF

Dataset 1 consists of AWCF features available in Grammarly's free version to collect data on students' experiences in the writing class. For each sampled text, the researcher initiated an editing process using the AI-powered tool, documented and coded it, according to the multimodal and pragmatic dimensions of the AWCF. To ensure reliability, intra-rater consistency was established by having the researcher re-code 100% of the data after a two-week interval. The two rounds of coding were then compared, showing a high level of consistency across categories. This ensured full internal consistency and minimised potential bias in the categorisation process, addressing a potential limitation of this exploratory study.

This procedure enabled the researcher to collect quantitative data (e.g., the initial and final scores obtained following AWCF and the total number of

recommendations) and qualitative information (the type of feedback provided verbally and through visual cues such as colours and images) (Table 2).

Text no.	Initial image and message	Number of words	Initial score	Final score	Total number of suggestions	Suggestions available in the free version	Suggestions available in the premium versions
1		136	78	96	9	8	1

Table 2
Multimodal grid.

To gain a deeper understanding of the pragmatic strategies employed by the AI-powered tool, feedback was further classified in terms of its illocutionary force (Table 3). The evaluation category refers to how AWCF performs the illocutionary function of assessment through scores, visual elements and written comments. The explanation category investigates how AWCF seeks to enhance understanding by offering clarifications, elaborations, or rationales, functioning pragmatically as assertive or informative illocutionary acts. The suggestion category refers to proposals for changes to the text that serve as directives. The encouragement category includes motivational feedback that performs an illocutionary function of motivating or supporting the learner, typically aligned with directive or expressive speech acts (Liu 2017; Weimei *et al.* 2024).

Text no.	Feedback item	Modality (visual, written, both)	Evaluation (positive, negative)	Explanation (explicit rule, example)	Suggestion	Encouragement
Text 1		Both	-	-	-	Present

Table 3
Table for multimodal and pragmatic analysis of AWCF.

2.2.2. Datasets 2 and 3: students' and teachers' insights on AWCF

Datasets 2 (student questionnaire) and 3 (teacher questionnaire) contain both quantitative and qualitative data on AWCF, gathered from the perspectives of students and teachers (Table 4). As the focus was on describing patterns in the data, responses to closed-ended questions were examined using descriptive statistics, such as frequencies, percentages, and means, without performing inferential statistical analyses. Accordingly, open-ended question

responses are analysed qualitatively using thematic analysis to identify recurring patterns and themes (Flick 2018). Drawing on research on digital classroom discourse, the thematic analysis of students' answers examines multimodal feedback from technical, socio-emotional, and personal perspectives (Chong 2019). While considering technical aspects, the analysis focuses on features and usability of AWCF including clarity and specificity of comments, as well as visual markers such as highlighting, colour coding. Socio-emotional aspects of feedback are examined through the tone conveyed in the comments (supportive versus critical, judgmental or lacking warmth) and their effect on students. Finally, personal aspects of feedback are considered by analysing students' needs, beliefs, and individual preferences regarding feedback (Chong 2019).

Teachers' responses were examined in terms of their feedback habits, attitudes toward multimodality, and strategies for pragmatic effectiveness, which were interpreted as manifestations of their pragmatic competence contributing to their CIC (Classroom Interactional Competence). However, given the focus of the study and the teachers' limited familiarity with AI, the teacher questionnaires yielded only limited data on their experiences with such tools, shifting the discussion more toward their feedback practices. Consequently, while the teachers' insights are valuable, they occupy a relatively limited space in the data presentation and discussion. This is not considered a limitation of the study, as the primary aim was to explore broader patterns of feedback practices rather than to assess teachers' direct engagement with AI tools.

Dataset	Source of information	Description	Focus	Theoretical reference(s)
Dataset 1	AWCF	Feedback instances on student writing	Evaluation, explanation, suggestion and encouragement	Pragmatic and multimodal discourse analysis
Dataset 2	Student questionnaire	Perceptions of feedback tone, helpfulness, motivation	Technical, socio-emotional, and personal perspectives	Thematic analysis; Descriptive quantitative analysis
Dataset 3	Teacher questionnaire	Views on pedagogical value, alignment with classroom feedback	Technical, socio-emotional, and personal perspectives	Thematic analysis; Descriptive quantitative analysis

Table 4
Data analysis overview.

Despite being a small-scale study, the data collected for this study are reliable, valid and ethical. Reliability was ensured through the use of standardized instructions and procedures across contexts. For qualitative data, intra-rater reliability was established through a full re-coding of all data after

a two-week interval. Validity was supported through clear coding protocols and alignment with the study objectives. Moreover, to gain more comprehensive insights and reduce possible personal bias, data from the three datasets are triangulated. Ethical standards were observed throughout the study. Participants were informed of the purpose and procedures of the research and provided informed consent prior to participation; for minors, parental or guardian consent was also secured. Informants were briefed at the beginning and debriefed at the end of the study to ensure understanding and voluntary participation. Participants were assigned codes and any identifying information was removed to ensure anonymity and confidentiality in the following section. Participation was entirely voluntary and free, and participants were informed that they could withdraw at any time without penalty. Anonymity and confidentiality were ensured by assigning codes and removing all identifying information from the data.

3. Findings

3.1. Pre-editing phase: Encouragement starts before writing

After opening Grammarly's interface and clicking on the "New document" button, users land on a blank writing space where they can enter the text. This is where the pre-editing phase occurs, as the AI tool greets the user with a multimodal message. Data reported in Table 5 show that encouragement is the key illocutionary force of speech acts in the pre-editing phase. In 100% of the cases, the software encourages users to start the writing process with multimodal locutionary acts consisting of comic-style and playful icons such as a paper plane, or a quill accompanied by a motivational message in bold font like *you got this* and an explanatory message in plain font (*suggestions will appear here*). These speech acts can be classified into three types: 1) expressives (e.g., *Rooting for you*); 2) directives (e.g., *Let's get to work*) 3) assertives (*Nothing to see yet*). The message *Suggestions will appear here*, which is displayed in every initial multimodal encouragement, is a multifunctional speech act. It informs the user (assertive) and guides their attention to a specific area (directive) but it reassures them about the help they will receive (expressive). The cartoonish elements and informal expressions of these speech acts convey a light-hearted and approachable tone also in the case of assertives and directives. For example, in the case of the suggestion *Start writing to make the magic happen*, the directive tone is mitigated by the reference to "the magic", which focuses on the positive, possibly transformative outcomes that will result from that action suggested.

Locutionary act		Type
Initial icon	Initial message	
 Rooting for you. Suggestions will appear here. Icon 1	Rooting for you	Expressive
 You got this. Suggestions will appear here. Icon 2	You got this	Expressive/indirect directive
 No pressure. Suggestions will appear here. Icon 3	No pressure	Expressive/indirect directive
 Let's get to work. Suggestions will appear here. Icon 4	Let's get to work	Directive
 Time to dig in. Suggestions will appear here. Icon 5	Time to dig in	Directive
 Start writing to make the magic happen. Suggestions will appear here. Icon 6	Start writing to make the magic happen	Directive
 Nothing to see yet. Suggestions will appear here. Icon 7	Nothing to see yet	Assertive
 The first word is the hardest. Suggestions will appear here. Icon 8	The first word is the hardest	Assertive
 All the world's a page. Suggestions will appear here. Icon 9	All the world's a page	(Metaphorical) assertive

Table 5
Overview of visual encouragement at the beginning of the editing process.

Students' questionnaire responses (Dataset 2) confirm the socio-emotional value of multimodal messages received in the pre-editing phase (Figure 1), which ease emotional tension and foster a supportive communicative environment.

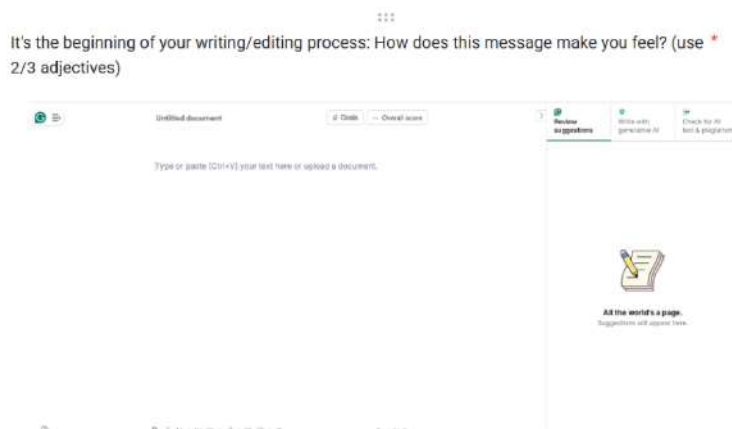


Figure 1

Example of a visual cue included by the researcher in the student questionnaire.

83% appreciate this page as it makes them feel *confident, encouraged, ready to write, motivated, free, relaxed/calm, taken care of* and *hopeful*. It reassures and empowers users to know that constructive feedback will be offered, and they will have the opportunity to enhance their writing (Comment 1).

(1) It makes me feel secure, because by this message i know that if i will get something wrong, the improvement/suggestion will appear here.²

Only a minority of respondents do not feel positively impacted by this multimodal message (10%) or reported negative emotions such as *confusion* or *sense of emptiness* (7%) (Comment 2).

(2) It makes me feel empty because I don't know where to start.

In addition to the AI-related risks reported in UNESCO (2021), such as over-reliance, misinterpretation of feedback, and ethical considerations, student feedback in this study revealed another concern. Consistent with Fan (2023), our findings indicate that AWCF may lead to experiencing emptiness, which may reduce motivation and foster a sense of inadequacy. AI should be used under teacher supervision or after student training sessions to address AI-

² While all identifying information has been removed to maintain anonymity, students' open-ended comments are reported verbatim to preserve the authenticity of participants' voices. To preserve authenticity (Morgan 2022), they have not been edited and may contain errors.

related risks, increase students' motivation and improve their attitude towards writing. However, Dataset 3 indicates that this is not a common teaching strategy among our informants. Eight teachers indicated they are not familiar with AI-powered editing tools such as Grammarly, while two reported using these tools to edit their own texts rather than as a teaching tool. When asked about strategies to encourage and motivate learners to write, they mentioned practices such as choosing relatable topics (N=7), collaborative work (N=5), and praise and evaluation (N=3). While teachers' feedback is varied, it does not appear to fully utilise the illocutionary potential of visual cues and supportive language to effectively accomplish the speech act of encouragement. This indicates that aspects of CIC, particularly those related to the strategic use of multimodal and affective resources, may not be fully realised in our informants' feedback practices. Moreover, this misaligns with students' appraisal of playful multimodal messages as fostering a nurturing environment for student motivation and empowerment, which is highly beneficial for enhancing learners' writing skills (Ducca 2014).

To delimit the scope of the present study, differences in the impact of written and visual components of multimodal messages were not analysed. While such distinctions may yield finer-grained insights, they fall outside the methodological focus of this paper and will be addressed in future research.

3.2. The editing phase: AWCF to improve student writing

3.2.1. AWCF evaluation and its perlocutionary impact

After pasting their texts through the AI-powered feedback generator, student participants receive multimodal feedback, including quantitative and qualitative insights. As the page layout shows, the focus shifts from the writer to the text (Figure 2). In this stage, AWCF performs the illocutionary function of evaluation and explanation while also providing suggestions. The former is expressed numerically, which in Searle's taxonomy (1979) can be classified as an assertive, since it presents the score as an objective assessment of textual quality (Hyland and Hyland 2001).

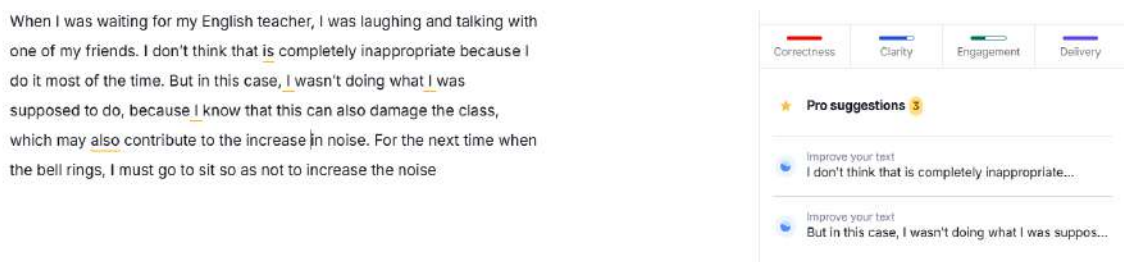


Figure 2
Example of editing page.

Despite its assertive tone, students clearly prefer numerical feedback to other types of feedback. Most of them rank it as the most useful and provide various reasons, ranging from practical aspects to the impact on them as learners. As shown in Comment 3, the score is *clear* and *easy* to understand; it *immediately* gives an idea of their level, the quality of their writing and their performance; it is the *closest thing to what a teacher would do*. AWCF makes them feel *focused* and *proud* and *helps them improve*. Moreover, as reported in Comment 4, numbers are straightforward, whereas the softening tone of some suggestions may lack clarity (Comment 4).

(3) I think it (*numerical feedback*)'s important, though not essential. It's important to understand your level roughly, but not essential because it doesn't fully determine your abilities. Anyway, it's the closest thing to what a teacher would do.

(4) I think it (*numerical feedback*) helps me more than just some simple suggestions because it helps me better understand and actually aim for 100.

This aligns with concerns raised in a previous study that shows how learners may find it difficult to gain a clear idea of their mistakes and the quality of their writing because of the softening strategies used to convey feedback (Hyland and Hyland 2001). However, a few comments point out critical aspects of numerical feedback. One student reports *negative feelings* as a consequence of the score received, whereas two other students think that it makes *you feel [like] a number* and *fails to acknowledge learners' general abilities*. In Comment 5, one person expresses confusion and disbelief after consistently receiving lower scores than expected, without understanding the underlying criteria of evaluation.

(5) I don't know, maybe I don't understand why my rating is like this and which criteria AI used to rank it.

Another aspect to consider regarding the perlocutionary effect of AWCF is personalisation: students appreciate Grammarly's interactive score, which updates dynamically as they revise their text. Comment 6 shows that although scores are primarily assertives, they may simulate or carry expressive force when personalised and increase motivation.

(6) The score changes and it helps me understand my progress I loved how the feedbacks were interactive and helped me reach a perfect score. I believe that Ai is a powerful tool and this is a good way to use it to improve yourself.

While students may find AWCF straightforward, it is important to caution them that this tool can create misleading expectations about how their writing

will be assessed, as it may not evaluate task requirements in the same way a teacher's expert assessment does. Table 6 displays the average trends in text length, initial score attributed by the AI-powered tool, and the total number of suggestions provided.

Word count range	Number of texts	Average Grammarly score	Average Number of Suggestions
Under 100 words	5 texts	74	12
100-200 words	8 texts	60	16
Over 200 words	7 texts	59	21

Table 6

Trends of average text length, average scores and suggestions provided by AWCF.

Shorter texts, particularly those under 100 words, achieve an average score of about 74 points. Texts containing 100-200 words average 60.8 points, whereas those over 200 words average approximately 59 points. This suggests that AWCF assesses shorter texts more favourably, this may be because longer texts provide more opportunities for grammatical or stylistic issues to appear. In line with this trend, longer texts receive more suggestions for improvement than shorter ones, which may give the impression that they are more inaccurate. This type of AWCF does not consider text length as a relevant feature of writing quality, which on the contrary may be relevant in task evaluation. In this regard, a potential concern arises that AWCF may mislead students who have not fully met the task's length requirement but still receive a high score with few suggestions for improvement, leading them to believe that they have successfully completed the task. Consequently, without teacher guidance, AWCF could have a limited perlocutionary effect and negatively impact students' learning.

3.2.2. *The perlocutionary force of multimodal suggestions for improvement*

AWCF integrates numerical scores with a combination of multimodal directive and assertive illocutionary acts to provide suggestions and explanations. While processing their texts, users receive an overview of the total number of suggestions for improvement, categorised according to different aspects (*correctness, clarity, engagement and delivery*) (Figure 3). This analysis focuses on the correctness category, specifically addressing red-underlined inaccuracies in AWCF for which automated suggestions are provided. Features related to clarity, delivery, and engagement are available only in the programme's premium version, which was beyond the scope of this study.

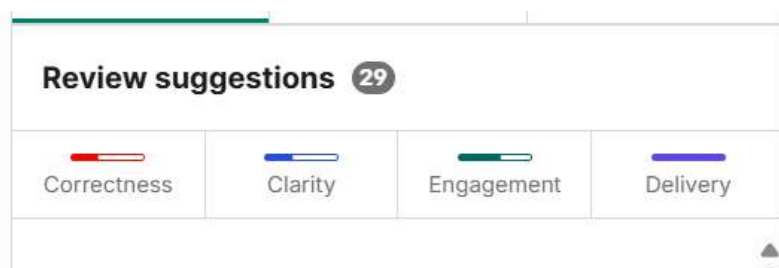


Figure 3
Overview of suggestions in AWCF.

Suggestions regarding correctness function as corrective directives that address clear-cut rules (e.g., grammar, agreement, typos) and require low cognitive demand. The combination of the visual mode of underlining inaccuracies in different colours (Figure 4) with clickable free suggestions for improvement in the sidebar (Figure 5) can be seen as a multimodal directive.

I was standing but near to my desk. I realize we were disturbing nearby classes and were a bit loud.

Figure 4
Example of free suggestion for improvement.

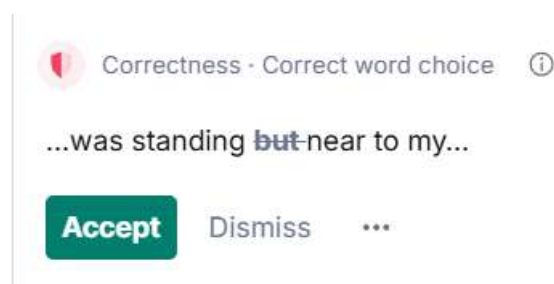


Figure 5
Example of text with visual cues.

However, the possibility to accept or dismiss the suggestion mitigates the directive tone of the suggestion (e.g., *fix capitalisation*, *check wording*, *add a comma*, etc.) while enhancing student agency. Student respondents do not mention the directive tone in their comments; instead, they focus on the positive aspects of multimodal AWCF, which they see as systematic, immediate, and *always available*. (e.g., Comments 7). Moreover, they appreciate the quality of the feedback they receive (Comment 8).

(7) The thing that I like the most is the fact that I have the feedback immediately

(8) (I like) The difference between the suggestion. They are very clear and

easy

This is confirmed by students' suggestions to improve teacher feedback, which recall AWCF. They think that feedback should include numerical information, written and oral (as in Comment 9). Moreover, students would value receiving feedback with colour categorisation (Comment 10) or a list with areas needing improvement (Comment 11).

(9) Feedback should be at the same time

- 1- Numerical, the grade to make a conclusion of the test/essay
- 2- Written, explanations of the mistakes
- 3- Oral, a further explanation of how to get over the mistakes

(10) (I would suggest) using different colours for grammar, vocabulary etc

(11) I would suggest to put at the end of the writing a list of the aspects you should improve, such as Grammar, Vocabulary etc... it gives a clear idea to the student of the weak points of his writing

In this regard, a discrepancy emerges between students' expectations for more varied and multimodal feedback and the extent of teachers' CIC in delivering it. Only two teacher respondents utilise coloured markings to indicate error severity, with most preferring to underline mistakes in red. Besides assigning numerical scores, these respondents frequently include written comments on students' work or offer direct sentence revisions to model appropriate language use. Seven respondents report using rubrics presented before tests to clarify evaluation criteria, help students understand how their grades are determined, and identify weaknesses and areas for improvement. However, while beneficial, teachers report that students often perceive this practice as summative rather than formative, limiting its potential to support learning. Interestingly, both student and teacher respondents are generally satisfied with teacher-generated feedback: none of the students express negative opinions about the feedback they receive, and the teacher participants report having never received complaints. Nonetheless, all our teacher respondents describe providing feedback as a *challenging, demanding* and *time-consuming* task, even though they acknowledge its fundamental role in learning. Providing regular feedback beyond formal assessment and offering personalised advice appear to be the aspects of feedback they struggle with most. Despite these challenges, none of our respondents has ever considered integrating AWCF into their teaching practice, indicating that the possibility of employing AI as a feedback partner has not yet been explored by the participants.

3.2.3. Explanations in AWCF and their perlocutionary effect

AWCF provides explanatory utterances, but only when requested by the user. If they require further clarification, users can click the information icon (① in Figure 5), which opens a window with grammatical explanations and sample sentences (Figure 6). The examples show that the explanations in AWCF combine directives (use this version, correct, improve your text) and assertives (Conjunctions are words like but, etc.).

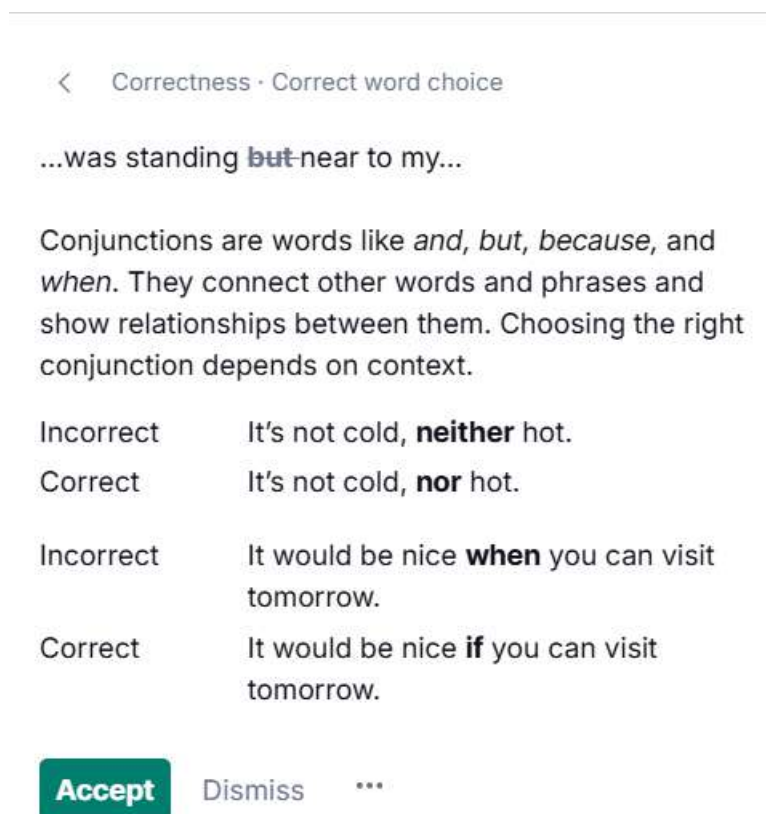


Figure 6
Example of window with grammatical explanations.

Based on our teacher respondents' experience, explanations are typically given orally when returning tests; however, they are not formally included as part of the provided feedback. Consequently, students who require additional clarification either approach the teacher individually or consult their textbooks or online resources. Teachers note that many students hesitate to request explanations in front of their peers, which often results in lingering questions and an ongoing need for clarification. This pattern points to areas of teacher CIC that could be strengthened by employing AWCF, particularly in fostering an environment that encourages open interaction and responsive clarification during whole-class exchanges. This need is confirmed by our student respondents, who value the opportunity to actively interact with

AWCF, either by immediately accepting suggestions or by requesting additional clarification, thereby retaining final authority over text editing (Comment 12).

(12) Not only that, the Ai didn't do everything by itself but, instead, it helped the writer in the process.

This finding demonstrates that students actively engage with AWCF, which contrasts with a common concern in the literature: the risk that students may become overly reliant on AI as a substitute for their own effort (Koltovskaia 2020). However, it also highlights a potential drawback of AWCF, as illustrated in the passage from Figures 4 and 5 to Figure 6. On the one hand, Figures 4 and 5 further illustrate how AWCF suggestions are personalised, referring point by point to words in the text. In the sentence *I was standing but near to my desk*, the word *but* is correctly underlined in red and removed in the suggested correction. On the other hand, the explanation in Figure 6 may mislead students, as it emphasises conjunction choice instead of indicating that a conjunction is unnecessary in this case. This discrepancy shows that although AWCF is generally helpful, it may provide unrelated explanations that confuse and frustrate students. The potential shortcomings of AWCF may impact its perlocutionary effect, making it important for both teachers and students to be attentive and aware.

3.3. Editing completion and praise through AWCF

Even with the free version of Grammarly, students showed a marked improvement in the quality of their written production, as evidenced by the difference between pre-editing and final scores. As shown in Figure 7, all texts increased from the initial to the final score. For example, Text 9 increased by 3%, from an initial score of 95 to a final score of 98, while Text 4 increased by 176%, from an initial score of 34 to a final score of 94.³

³ Percentage increases were calculated as $[(\text{final score} - \text{initial score}) \div \text{initial score} \times 100]$.

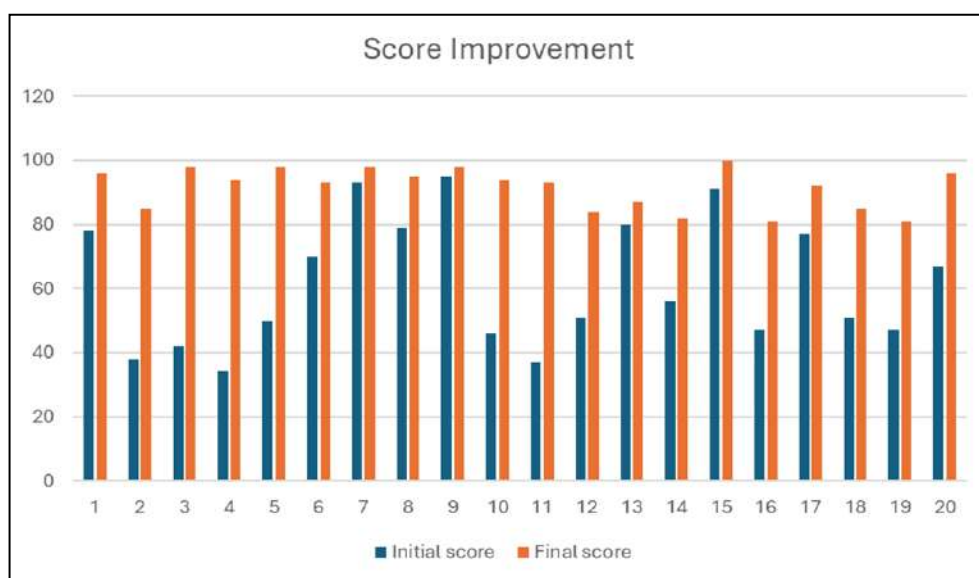


Figure 7
Overview of score improvements.

This suggests that the editing or intervention process consistently had a positive impact on text quality and perceived improvement. Even if improved text quality does not automatically translate into better writing skills, it seems relevant that AWCF can help students feel empowered. This could positively influence students' self-esteem, self-efficacy, and overall learning, particularly among less proficient learners. Texts with lower initial scores (e.g., Texts 2, 3, 4, 10, 11), which had more room for improvement, showed significant increases, often exceeding a 100% increase. By contrast, texts that started with already high scores (e.g., Texts 7, 9, 13, 15) showed small percentage increases, typically under 10%. This suggests that AWCF can make a difference, especially for students who need to improve their writing skills and need encouragement and scaffolding.

In this regard, the use of praise and encouragement is particularly important. Positive feedback can reinforce effort, boost confidence, and help students engage more actively with the learning process (Ducca 2014). For most students, the editing process end with a final evaluation; notably, 85% of respondents achieve scores exceeding 90%. Dataset 2 shows that only one student attains a perfect score of 100%. The multimodal praise obtained for scoring 100 is included in the student questionnaire to gather students' insights into the end of the editing process (Figure 8⁴).

⁴ This figure refers to the end of the editing process performed by the researcher to compile Dataset 1 resulting in a 100% score. To re-create the same conditions in which the students worked, the researcher limited herself to following the suggestions provided in Grammarly's free version.

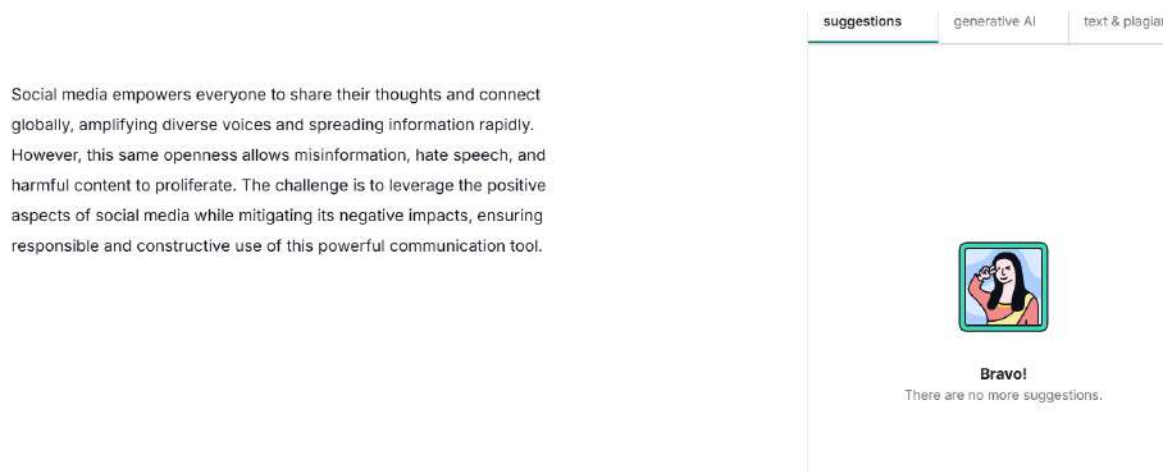


Figure 8
Example of multimodal feedback for successful editing.

In addition to evaluating the text, AWCF performs a multimodal expressive illocutionary act combining process praise (the reference to Mona Lisa values the writing product as a masterpiece) with an expressive form of person praise (*Bravo*) and assertive comment (*there are no more suggestions*). Interestingly, this multimodal speech act overcomes the traditional separation between praise targeting the person (e.g., *you are clever*) and praise focusing on the process (e.g., *you worked hard*), by integrating affective and instructional dimensions into a single communicative act (Skipper, Douglas 2012). It acknowledges the learner as a person (*Bravo*) while praising the outcome of their effort, with a score and a comment about the lack of further suggestions. This seems to reinforce both self-efficacy and a growth-oriented mindset. Moreover, the message displayed in Figure 8 is gained after following the suggestions which impact the personal score; this may lead learners to perceive it as personalised and well-deserved praise. This seems particularly valuable given the benefits of personalised feedback in contrast with general feedback, perceived as insincere (Skipper and Douglas 2012). The survey results indicate a positive perlocutionary effect of this speech act. Students respond favourably to the use of a numerical score (100%) combined with an icon and a positive message. Participants describe the message as clear (Comment 13), and one respondent indicates that AWCF has a positive effect by conveying a sense of achievement (Comment 14).

(13) It's pretty clear it tells me that I've done a good job and my text was clear and perfect.

(14) It makes me feel completed when I score 100.

When asked to define how rewarding the multimodal message in Figure 8 is, 83 % of the students (albeit in different percentages) report it as *rewarding*.

Although the “100 score” is an interesting combination of affective and instructional dimensions, it seems to carry potential risks. Because the praise is directly tied to the score, learners may internalise it as an evaluation of themselves rather than their performance. Moreover, the perlocutionary effect of such messages could create differences between high- and low-scoring students, thereby mitigating the positive impact of AWCF. As already pointed out in 3.2.1 and 3.2.3, although effective, AWCF may be problematic when used by students without adequate training.

Regarding the potential benefits of multimodal praise, students and teachers express differing views, reflecting a gap in their perceptions of its value. While students value the delivery of praise in AWCF, teachers demonstrate greater caution, suggesting limited confidence in the pedagogical potential of multimodal feedback and its integration within their classroom interactional competence. Although four teacher respondents acknowledge that the message in Figure 8 can be motivating, none would use it with their students for various reasons. This type of feedback is *suitable for younger students* (N=8), *it adds little value to feedback* (N=4) or can even feel *impersonal and automatically attributed* (N=4). Teachers prefer expressing their praise verbally in written (N=8) or oral mode (N=2). Eight of them use positive, process-oriented feedback (well done! Good job!) and prefer to praise the writing process rather than the final product. They adapt their praise to the students’ profiles to make it more effective (Comment 15).

(15) If I know a student has a low self-esteem I try to make the mistakes clear but at the same time I focus on what he/he did well to encourage them

4. Conclusions, limitations and implications

Given the rapid and widespread expansion of AI, educators have an important opportunity to lead the conversation on how to meaningfully integrate it into the teaching and learning process. This case study explores AI-powered multimodal feedback as a potential extended pragmatic partner that can integrate teacher CIC to foster students’ writing proficiency. To this end, the study integrates data from a pragmatic analysis of Grammarly-generated multimodal feedback on student texts with insights from teachers and students.

The analysis indicates that AWCF is capable of performing illocutionary acts related to feedback, such as providing evaluations, suggestions, explanations, and encouragement. Interestingly, encouragement comes first when editing texts with the AI-powered tool. As discussed in 3.1, visual cues such as cartoonish icons and encouraging messages are used to create a supportive and motivating environment, which is appreciated by

students. Encouragement in AWCF uses expressives, directives and assertives to motivate learners. The playful tone set by the combination of icons and informal language mitigates the force of expressives and directives, making them less face-threatening, which students appreciate. In comparison, teachers tend to view this strategy as particularly suitable for younger learners, which may reflect the current focus of their CIC on conventional feedback practices rather than on the full illocutionary potential of multimodal messages that combine visual cues and supportive language. For older students, our teacher informants tend to use alternative approaches that focus on extrinsic and intrinsic motivation, rather than relying on visual elements or a playful tone. Although this study does not recommend that teachers adopt playful elements such as cartoonish icons or emojis, it aims to highlight the role of multimodal encouragement as a powerful perlocutionary act that motivates learners and enhances their sense of achievement. More specifically, students' appreciation of this aspect of AWCF suggests that incorporating AI-powered feedback into teaching practices may broaden teachers' pragmatic competence, even when they do not directly use these multimodal messages. This could help teachers create a pressure-free environment in their EFL writing classes, allowing students to enjoy the experience and appreciate the learning process, thus overcoming the limits of an assessment-focused approach.

The limited familiarity of our teacher respondents with AWCF restricts the extent of analysis dedicated to their insights. Nevertheless, teachers' answers provide valuable qualitative perspectives on feedback practices. Their comments highlight process-oriented praise, personalisation, and the balance between encouragement and critique, which complements the AWCF findings by illustrating how human feedback supports motivation and creativity. Thus, although discussed concisely, teachers' perspectives remain integral to understanding the broader pedagogical value of feedback in writing classes. Our findings suggest that students could benefit from AWCF, as it can address needs that are not always fully met by traditional feedback. At the same time, AWCF could serve as a pragmatic extension for teachers, integrating and enhancing their feedback practices while alleviating some of their workload and enhancing their CIC. AWCF analysis in 3.2 and 3.3 reveals that AI-powered tools systematically provide feedback while performing illocutionary acts like evaluation, suggestion and praise, which are typically separated or less prominent in teacher feedback. Due to time constraints and heavy workloads, teachers frequently face challenges in providing personalised, timely feedback independent of assessment activities. Research shows that this is not beneficial to learning as it shifts the focus from the process to the result, thus encouraging an assessment-oriented perspective. Additionally, recommendations for improvement and detailed explanations are typically delivered orally, as preparing such feedback is a

time-intensive responsibility for educators. By contrast, AI-powered tools provide multimodal interactive feedback, including evaluation (numerical scores and visual cues), suggestions for improvement and explanations, while giving students the possibility to monitor their progress. Moreover, students report that AWCF's numerical scores are clear and motivating, encouraging them to refine their writing in order to achieve a higher score. From a pragmatic perspective, AWCF can integrate teacher-assigned scores—assertive acts that may be face-threatening—while mitigating their potential discouraging effects through supportive, multimodal feedback. Therefore, this study argues that AWCF should be integrated into teaching strategies, as it can both expand and empower teacher feedback, thus enhancing instructors' pragmatic skills to maximise learning potential.

Our learner informants value AI-generated feedback as systematic, easily accessible, and partially personalised. The findings show that it meets students' needs from a technical, socio-emotional and personal perspective. Moreover, the possibility to monitor their progress is, *per se*, empowering for students and transforms the editing process into a rewarding experience. This indicates that AWCF can also serve as a standalone tool for students' self-study and autonomous writing development, though it still requires appropriate teacher supervision and student training to ensure effective use. Students should be aware that, while Grammarly can be effective and helpful, it is not designed as an educational tool. As such, it does not guarantee academic success and does not prioritise student understanding and development in the way teachers do. The analysis points out possible AWCF's shortcomings, which lessen its perlocutionary value. For example, AWCF may be inaccurate, offer unrelated suggestions, and mislead students in evaluating their text quality, as it cannot be tailored to specific requirements, such as text length. Moreover, AWCF's lack of context awareness and longitudinal tracking limits its illocutionary relevance and may weaken its perlocutionary effect on student motivation and understanding.

The present study has several limitations. First, it is an exploratory small-scale study, and its results cannot be generalised. The participation of only ten teachers may limit the range of experiences and perspectives represented in this study. Consequently, the findings should be regarded as illustrative examples rather than broadly representative conclusions. Moreover, the student participants reportedly had proficiency levels ranging from B1 to C1, which limits the applicability of the findings to less proficient learners, who may require more scaffolding. Consequently, aspects such as the difficulties arising from the fact that AWCF is delivered in English were not considered in the analysis or discussion. Second, the teacher informants had limited experience with AWCF, which prevented a detailed point-by-point comparison between teachers' and students' insights. Third,

establishing a correlation between AWCF use and students' development of writing skills was neither feasible nor within the scope of this study.

Although this study cannot demonstrate that integrating AWCF into EFL writing instruction directly improves students' writing quality, it provides valuable insights into its potential pedagogical implications. Amid the ongoing debate over AWCF's impact on student writing, this study supports its use within a blended model of human and AI-generated feedback, emphasising that implementation should occur under teacher supervision or after student training. Students' appreciation of AWCF indicates that it can serve as a complement to teacher feedback, helping improve instructional effectiveness and further develop teacher CIC. In supporting students' editing through Grammarly, teachers can benefit from the affordances of AI-powered tools to expand and diversify the feedback they provide. At the same time, teachers should offer guidance to help students deal with inaccurate, unclear and confusing AWCF. In addition to that, they can integrate AI-provided feedback with their knowledge of individual student's situations and put it into a perspective that accounts for students' improvements over time. So, AWCF can serve as a significant element in teaching practices aimed at developing innovative teaching methods to improve learner writing abilities. However, the findings also indicate that teachers may still be developing their own familiarity with AI for pedagogical purposes, highlighting a need for targeted professional development.

Given the rapid and ongoing technological advancements, further research is needed. Possible research strands could investigate the effects on other writing-related factors such as self-efficacy, self-regulated strategies, and factors that influence perception, use, or engagement with AWCF. It might be relevant to study how AI-powered tools can support learning based on students' age, proficiency level, and individual learning preferences or needs, including cognitive style, learning pace, and neurodiversity. Ultimately, integrating AI-powered multimodal feedback into EFL writing classrooms may strengthen teachers' professional pragmatic skills and contribute to more effective pedagogy. To maximise AI's potential as an extended pragmatic partner for teachers, the research directions identified in this study should be incorporated into teacher training programs, enabling educators to make informed, pedagogically sound use of AI tools.

Bionote: Antonella Giacosa (PhD) has been an EFL teacher in secondary education for over 20 years and has collaborated with the English Department at the University of Turin since 2017. She is an educational linguist with a strong interest in innovative, intercultural, and transdisciplinary approaches to English Language Teaching (ELT). She has served as a contract lecturer on the MA programme in English Language and Linguistics, with a focus on ELT. She is a member of research teams dedicated to curriculum

internationalization and innovation at both school and university level, in Italy and abroad. In 2024, she published her first monograph, *Effective Multimodal Interaction for Online and Hybrid Teaching: Data-driven Insights* (Palgrave Macmillan).

Author's address: antonellamaria.giacosa@unito.it

Disclosure statement: The author declares no conflicts of interest. This research was conducted independently and received no financial support or affiliation with Grammarly or any related organisation.

References

- Ajjawi R., Kent F., Broadbent J., Tai J.H.M., Bearman M. and Boud D. 2022, *Feedback That Works: A Realist Review of Feedback Interventions for Written Tasks*, in “Studies in Higher Education” 47 [7], pp. 1343-1356.
- Alimova S., Qodirova T., Sobirova D., Turdimamatova L. and Hafiziddinova N. 2025, *Pragmatics of Teacher’s Speech in the Classroom*, in “Modern Education and Development” 25 [6], pp. 55-60.
- Bayat M., Shirvan M.E. and Barabadi E. 2020, *A Teacher’s Multimodal Corrective Feedback: Exploring Its Enjoyment Building Capacity*, in “Polish Psychological Bulletin” 51 [2], pp. 71-88.
- Carless D. and Winstone N. 2020, *Teacher Feedback Literacy and Its Interplay with Student Feedback Literacy*, in “Teaching in Higher Education” 28 [1], pp. 150-163.
- Chong S.W. 2019, *College Students’ Perception of e-Feedback: A Grounded Theory Perspective*, in “Assessment & Evaluation in Higher Education” 44 [7], pp. 1090-1105.
- Creswell J.W. and Creswell J.D. 2018, *Research Design. Qualitative, Quantitative and Mixed-Method Approaches*, 5th ed., Sage Publications, Los Angeles (CA).
- Cunningham K.J. 2019, *Student Perceptions and Use of Technology-Mediated Text and Screencast Feedback in ESL Writing*, in “Computers and Composition” 52, pp. 222-241.
- Darong H. and Guna S. 2023, *Corrective Feedback: Pragmatic Exposures in EFL Classroom Interactions*, in “International Journal of Education and Learning” 5 [1], pp. 14-22.
- Ducca J.D. 2014, *Positive Oral Encouragement in the EFL Classroom: A Case Study Through Action Research*, in “Revista de Linguas Modernas” 21, pp. 325-346.
- Ekin M. and Damar E. 2013, *Pragmatic Awareness of EFL Teacher Trainees and Their Reflections on Pragmatic Practices*, in “ELT Research Journal” 2 [4], pp. 176-190.
- El-Dakhs D.A., Ambreen F., Zaheer M. and Gusarova Y. 2019, *A Pragmatic Analysis of the Speech Act of Criticizing in University Teacher-Student Talk: The Case of English as a Lingua Franca*, in “Pragmatics” 29 [4], pp. 493-520.
- Escalante J., Pack A. and Barrett A. 2023, *AI-Generated Feedback on Writing: Insights into Efficacy and ENL Student Preference*, in “International Journal of Educational Technology in Higher Education” 20, pp. 1-20.
- Fan N. 2023, *Exploring the Effects of Automated Written Corrective Feedback on EFL Students’ Writing Quality: A Mixed-Methods Study*, in “SAGE Open” 13 [2], pp. 1-17.
- Farnia M., Hemmati F., Rabbani R. and Mirsayafi S. 2025, *The Impact of Multimodal Feedback on Writing Complexity, Accuracy, and Fluency of Iranian EFL Learners in Flipped and Traditional Classrooms*, in “Journal of Applied Linguistics and Applied Literature: Dynamics and Advances” 13 [1], pp. 109-130.
- Flick U. 2018, *An Introduction to Qualitative Research*, 6th ed., Sage, London.
- Herring S. 2019, *Grammar and Electronic Communication*, in Chapelle C.A. (ed.), *The Encyclopedia of Applied Linguistics*. <https://onlinelibrary.wiley.com/doi/abs/10.1002/9781405198431.wbeal0466.pub2> (17.11.2025).
- Huawei S. and Aryadoust V. 2024, *A Systematic Review of AI-Based Automated Written Feedback Research*, in “ReCALL” 36 [2], pp. 187-209.

- Hyland F. and Hyland K. 2001, *Sugaring the Pill: Praise and Criticizing in Written Feedback*, in "Journal of Second Language Writing" 10, pp. 185-212.
- Ishihara N. 2011, *Co-Constructing Pragmatic Awareness: Instructional Pragmatics in EFL Teacher Development in Japan*, in "TESL-EJ" 15 [2], pp. 1-16.
- Ismail S.M., Rahul D.R., Patra I. and Rezvani E. 2022, *Formative vs. Summative Assessment: Impacts on Academic Motivation, Attitude Toward Learning, Test Anxiety, and Self-Regulation Skill*, in "Language Testing in Asia" 12 [1], pp. 1-23.
- Jiang L., Lee I. and Yu S. 2024, *Conceptualizing Multimodal Feedback Literacy for L2 Writing Teachers in the Digital Age*, in "International Journal of Applied Linguistics" 34 [4], pp. 1479-1496.
- Kerr P. 2020, *Giving Feedback to Language Learners*, Cambridge University Press, Cambridge.
- Koltovskaia S. 2020, *Student Engagement with Automated Written Corrective Feedback (AWCF) Provided by Grammarly: A Multiple Case Study*, in "Assessing Writing" 44, no. 100450.
- Kress G.R. and van Leeuwen T. 2020, *Reading Images: The Grammar of Visual Design*, 3rd ed., Routledge, London.
- Liu Y., Liu X. and Zhang Y. 2021, *A Pragmatic Study on Teachers' Feedback in EFL Classroom in China*, in "English Language Teaching" 14 [5], pp. 109-123.
- Moorhouse B.L. 2024, *Beginning and First-Year Language Teachers' Readiness for the Generative AI Age*, in "Computers and Education: Artificial Intelligence" 6, no. 100201.
- Moorhouse B.L., Li Y. and Walsh S. 2021, *E-Classroom Interactional Competencies: Mediating and Assisting Language Learning During Synchronous Online Lessons*, in "RELC Journal" 54 [1], pp. 114-128.
- Morgan H. 2022, *Conducting a Qualitative Document Analysis*, in "The Qualitative Report" 27 [1], pp. 64-77.
- Nava A. and Pedrazzini L. 2018, *Second Language Acquisition in Action: Principles from Practice*, Bloomsbury, London.
- Ranalli J. and Yamashita T. 2022, *Automated Written Corrective Feedback: Error-Correction Performance and Timing of Delivery*, in "Language Learning & Technology" 26 [1], pp. 1-25.
- Searle J.R. 1979, *Expression and Meaning: Studies in the Theory of Speech Acts*, Cambridge University Press, Cambridge.
- Skipper Y. and Douglas K. 2012, *Is No Praise Good Praise? Effects of Positive Feedback on Children's and University Students' Responses to Subsequent Failures*, in "British Journal of Educational Psychology" 82, pp. 327-339.
- Steiss J., Tate T., Graham S., Cruz J., Hebert M., Wang J., Moon Y., Tseng W., Warshauer M. and Booth Olson C. 2024, *Comparing the Quality of Human and ChatGPT Feedback of Students' Writing*, in "Learning and Instruction" 91, no. 101894.
- UNESCO 2021, *Recommendation on the Ethics of Artificial Intelligence*, <https://unesdoc.unesco.org/ark:/48223/pf0000381137> (10.07.2025).
- van den Berg G. 2025, *Teachers' Experiences of Using Artificial Intelligence From an Open Distance Learning Context: Successes, Challenges, and Strategies for Success*, in "Discover Education" 4, no. 192.
- Walsh S. 2014, *Developing Classroom Interactional Competence*, in "Language Issues: The ESOL Journal" 25 [1], pp. 4-8.
- Weimei L., Xiuwen C. and Lihe H. 2024, *Promoting Second Language Writing Through Technology-Driven Multimodal Text Feedback*, in "Innovation in Language

Learning and Teaching”, pp. 1-19.

Youn C., Salam A. and Rahman A. 2025, *AI-Driven Tools in Providing Feedback on Students' Writing*, in “International Journal of Research and Innovation in Social Science” 9, pp. 58-67.

REPRESENTATIONS OF HATE DISCOURSE

Interactions between Human Linguistic Pragmatics and AI Multimodal Pragmatics

RAFFAELE PIZZO
UNIVERSITY OF NAPLES L'ORIENTALE

Abstract – Considering the intertwining relationships between the Web 2.0 (O'Reilly 2005) and hate speech proliferation as an intrinsic peculiarity of contemporary societies (Balirano and Hughes 2020), in which online communities of belonging are created through users' interaction and engagement (Zappavigna 2014), this paper compares human linguistic conception of hate speech with its corresponding AI multimodal design. More specifically, it analyses how the concept of hate speech is linguistically realised by university students according to the social categories of age, body, gender, disability, ethnicity, and religion, as defined within the research framework. The subsequent phase of analysis focuses on the ensuing AI-generated images to understand how linguistic input provided was interpreted and visually elaborated on. The two phases respectively apply Gricean principles of conversational cooperation (Grice 1975) and Kress and van Leeuwen's multimodal theory (2020). Through this approach, this article explores the various articulations of hate speech as shaped by individual perceptions and student-machine interaction, thereby highlighting potential biases of both socio-cultural and computational nature. Given the pervasive role of AI programs in contemporary societies, discrepancies may shed light on how the machine conceives, reprocesses, and realises the intended concept in ways that diverge from human intention, thereby potentially stripping the user of the agency that fundamentally characterises human communication. In essence, it examines the extent to which AI technologies can bolster or contrast the spread of hate discourse.

Keywords: hate speech; pragmatics; pragmalinguistics; multimodality; AI.

*[...] hate speech online does not occur in a
virtual vacuum, its effects are dramatically
real for those individuals who are on the
receiving end.*

(G. Balirano & B. Hughes, "Homing in on
Hate: Critical Discourse Studies of Hate
Speech, Discrimination and Inequality in the
Digital Age", 2020, p. vii).

1. Hate speech online

The proliferation of hate speech in digital environments has emerged as a central concern within English language studies (see, amongst others, Balirano and Hughes 2020; Ben-David and Matamoros-Fernandez 2016; Burnap and Williams 2015; Esposito and KhosraviNik 2018; Hardaker and McGlashan 2016; Klein 2010; Potts 2015; Yamaguchi 2013), which have been tackling it from the most disparate research perspectives and topics. A common thread and a defining feature of today's digital landscape is the so-called Web 2.0 (O'Reilly 2005), whose emergence massively transformed the way users interact both with each other and with the network itself. The core shift induced by Web 2.0 lies in the transition from static, unidirectional interaction – where passive users merely receive information and data from the web – to a bi- and multi-directional model of interaction, where Internet users actively contribute to shaping the web through content creation and digital interpersonal communication. While this evolution has undoubtedly brought numerous improvements to contemporary society, it has also fostered the proliferation of hate speech online. This is often a consequence of polarisation between digital language communities holding opposing views on a specific topic – thereby intertwining the Ideational and Interpersonal metafunctions with the Appraisal theory principles (see Zappavigna 2014a, 2014b; Zappavigna and Martin 2018) – further exacerbated by the perceived anonymity that the web is thought to provide. Moreover, the rapid acquisition of digital skills has not been accompanied by a corresponding development in digital awareness, leaving the web as a virtually unregulated space where seemingly anything is permissible. Rather than being detached from tangible reality, contemporary platforms – ranging from social media and online multiplayer games to augmented reality – constitute a substantial part of modern life, in which individuals are increasingly more accustomed to human-machine interactions rather than human-human ones. This trend has been further intensified by the advent of Web 3.0 and Web3¹ (see, among others, Choudhury 2014; Kantara 2018), which are characterised by decentralisation, blockchain technologies, and AI systems. The latter are currently revolutionising several sectors of contemporary societies, showing that digital advancements cannot be confined solely to the virtual domain; rather, they have significant implications for human existence. On the one hand, AI models have enhanced and accelerated logistics, industrial production, and machine translation (albeit not without challenges); on the other, they may also have internalised behavioural biases, as they learn from

¹ Despite being similar, these two conceptualisations are marked by subtle differences, whose detailed discussion, however, falls outside the scope of the present study.

human (linguistic) input. Against this backdrop, the present study aims to identify the sources of cognitive bias and stereotypes in human-machine conversational interactions. More specifically, it aims to highlight how human hate speech is confirmed, mirrored, and even amplified in AI multimodal productions, thus revealing both conversational and computational biases. To delineate the scope of this investigation, the study is rooted in the PRIN ECHOES Project.

1.1. The PRIN ECHOES project and the current study

The PRIN ECHOES² project aims to investigate contemporary forms of online hate speech from a multimodal semiotic perspective, as understood by lower secondary school students. Drawing inspiration from this research framework, the present study adopts an inverse approach: rather than collecting students' passive feedback, it investigates their active engagement with the concept of hate speech. In other words, whereas in the original project, middle and high school students (11-19 years old) were asked to comment on pre-selected examples via a tailored survey, this study poses the initial question directly to university MA students (21-25 years old), thereby granting them greater expressive freedom and reducing external influence. To enable this methodological shift, the students' level of education was raised – both for practical reasons and to ensure participants were fully aware of the study's objectives and sufficiently familiar with the concept of multimodality. While further details on the research methods adopted are provided in Paragraph 2, the overarching aims of the study and its research questions are clarified below.

Through the students' active engagement with AI systems, this investigation first seeks to identify, via a pragmalinguistic analysis of the communicative exchanges, whether such systems have internalised any human biases. By comparing the human linguistic construction of hate speech with the resulting AI multimodal outputs in conversations, it then highlights potential discrepancies, similarities, and conversational implicatures from a multimodal pragmatic perspective. The goal is to examine whether the original linguistic input has been accurately transmediated into the visual mode, while also offering an analysis of AIs' own multimodal conceptualisation. Ultimately, this article aims to determine whether – and in

² The PRIN (*Progetto di Rilevante Interesse Nazionale* [Project of Relevant National Interest]) ECHOES – English Studies to Contrast Hate Online and Enhance Solidarity – project ran from 2023-25 under the supervision of the Principal Investigator, Prof. Balirano Giuseppe (University of Naples L'Orientale), involving researchers from the University of Naples L'Orientale, the University of Naples "Parthenope", the University of Campania "Luigi Vanvitelli", and the University of Messina.

what ways – AI technologies may either contribute to or counteract the spread of online hate speech. To address these research questions and achieve the overarching aim, the methodologies and methods outlined in the following section are adopted.

2. Methodologies and methods

The methodological framework of this study is grounded in the concepts of pragmatics (see, among others, Austin 1975; Brown and Levinson 1987; Goodwin 2013; Goffman 1967; Grice 1975; Leech 1983) and multimodality (Kress and van Leeuwen 2006; O'Halloran 2004; O'Toole 1990, 1994, 2004), with the aim of contributing to research developments and advancements in the field of multimodal pragmatics (see, among others, Dicerto 2018; Mubenga 2009; O'Halloran *et al.* 2014; Payrató 2017). The latter approach is informed by Halliday's Social Semiotic theory (1978) and Systemic Functional Linguistics (SFL henceforth) (2014), which conceptualise communication as arising from the interaction of multiple semiotic resources, including language, visual elements, gesture, and embodied action. Multimodal pragmatics extends traditional pragmatic analysis by incorporating the contextual and functional use of non-verbal and material resources alongside language to explore how meaning is co-constructed in communicative events. Drawing on research by O'Halloran *et al.* (2014), this perspective emphasises the role of different semiotic modes in achieving communicative and instructional goals, particularly in digitally mediated and collaborative environments. The methodology is grounded in the principle that meaning is made through the coordinated deployment of multimodal resources within specific social and cultural contexts. As such, this approach allows for a comprehensive analysis of interactional dynamics, including how human and AI participants use language and visual tools to negotiate meaning and complete tasks collaboratively. In order to apply the aforementioned methodology, the specific theoretical approaches adopted in this study are outlined below: Grice's principles of conversational pragmatics (1975) and Kress and van Leeuwen's multimodal theory (2006). Section 2.2, instead, is devoted to the methods employed for the data collection, organisation, and analysis.

2.1. Methodologies

2.1.1. Grice's conversational pragmatics

This analysis is grounded in Grice's theory (1975) of conversational pragmatics, particularly the Cooperative Principle, the Conversational

Maxims – Quantity, Quality, Relation, and Manner – and the Implicature.

According to Grice, effective communication relies on the assumption that participants are cooperating. He formalises this in the Cooperative Principle (CP henceforth), which guides speakers to make their contributions appropriate to the conversational context. In order to do that, participants have to respect four conversational maxims:

- Quantity: Be as informative as necessary.
- Quality: Be truthful.
- Relation: Be relevant.
- Manner: Be orderly and concise.

If one of the interlocutors does not observe one or more of these maxims, the CP is violated, thereby giving rise to a potential conversational implicature. According to Grice's theory of implicature, speakers often communicate more than what is explicitly stated. Implicatures can be distinguished between:

- Conventional implicatures (tied to specific expressions, e.g., “but”).
- Conversational implicatures (inferred through the maxims).

Speakers may thus flout maxims to deliberately generate implicatures (e.g., irony, metaphor), or violate them covertly (e.g., lying), both of which affect how meaning is constructed. This research specifically focuses on the second type. Moreover, Grice's framework provides the basis for identifying both standard conversational implicatures and more nuanced social implicatures emerging from participant-AI interactions. Grice's maxims were used to determine where and how implicatures arose in the conversations, especially in instances where student or AI responses appeared to flout, violate, or suspend one or more maxims. Particular attention was paid to the pragmatic inferences participants made when interpreting or negotiating meaning with the AI (and vice versa) on sensitive topics like hate speech. The presence of social implicatures – such as politeness strategies, stereotypes, or identity-related positioning – was also analysed with a view toward how social meaning is co-constructed in human-AI interactions.

2.1.2. Kress and van Leeuwen's multimodal theory

Kress and van Leeuwen's grammar (2006) provides a systematic framework for analysing the meaning-making potential of visual compositions, which are not conceived as transparent representations. Their social semiotic theory treats images as culturally situated texts that function through structured visual grammar analogous to linguistic syntax.

Their model is based on three main metafunctions adapted from Halliday's SFL (2014):

- Representational metafunction: examines how participants (people, objects, places) and processes are visually depicted.
- Interactive metafunction: explores the relationships between the image, its producer, and the viewer.
- Compositional metafunction: considers how visual elements are arranged to create coherence and hierarchy.

These metafunctions serve as the primary categories for coding and interpreting the visual data, enabling a multimodal analysis that accounts for both form and meaning. This model has proven useful in the present study for analysing the semiotic meaning of AI-generated images and for identifying relevant details regarding how the machine visually processes the concept of hate speech.

By combining the two methodologies across the two research stages, it was possible to implement a multimodal pragmatic approach. This methodology added an additional conceptual layer to the study: examining how the AIs considered for the study generated the requested images from the linguistic input provided by students. In doing so, the research simultaneously emphasises three key aspects: how students conceptually and linguistically construct the notion of hate speech; how AIs interpret and render hate speech visually; and how human-machine interaction is currently shaped at a conversational level, enabling the identification of conversational and social implicatures, cognitive and social biases, and shifts in meaning resulting from the intersemiotic translation from language to image.

2.2. Methods, data collection and analytical procedure

Following the methodological approach adopted in the PRIN ECHOES project, this study involved the creation of a survey, which was subsequently administered to university students from March to May 2025 in order to collect the necessary linguistic and multimodal data. The survey included: a general explanation of the project along with guidelines for interacting with AIs and generating the required visual content on hate speech; a section for uploading the conversation with the AI; a question prompting students to explain their choice and interpretation of a specific social category; and a demographic section including age, gender, ethnicity, religion, and native language. These last two sections proved particularly useful for making inferences regarding the linguistic and multimodal results obtained. The corpus was thus constructed and comprises 17 survey responses, each consisting of one human-machine interaction and one AI-generated image.

The analysis began with a pragmalinguistic examination of the conversations, applying Grice's (1975) theory to identify potential implicatures, both conversational and social. The aim was to determine whether all four conversational maxims were observed, and, if not, to explore

the reasons for their violation. Subsequently, a multimodal analysis of the AI-generated images was conducted, using Kress and van Leeuwen's (2006) visual grammar framework. A key focus of the research was the comparison between the linguistic input and the visual output, with the objective of understanding whether, when, and why discrepancies occurred. This comparative approach allowed for the inclusion of additional observations and comments in the analysis.

Paragraph 3 presents the results and their arrangement and proceeds with the analysis and discussion, examining each student individually.

3. Pragmalinguistic and multimodal analysis

The analysis section follows the structure outlined below: each student-AI interaction is analysed individually, starting with the linguistic examination and proceeding with the multimodal analysis. In order to ensure a smooth and coherent presentation of the findings, personal information (age, gender, ethnicity, religion, native language), the underlying motivations behind the students' choice to focus on a specific social category and the rationale for its particular interpretation, as well as the full transcripts of the student-AI conversations are provided in the annex. However, the first meaningful turns of each conversation are presented at the beginning of each subparagraph to enhance readability and to illustrate the ways in which each student initiated and oriented themselves toward the task. In contrast, the AI-generated images are included within each respective subsection to enhance readability and comprehension³. Students were given the freedom to select their preferred AI system; however, ChatGPT and Copilot were recommended. All students interacted with ChatGPT, with the exception of Student VII, who resorted to ReVe AI.

3.1. Student 1

- Hello, can you create me an image about religion connected with hate speech?
- Sure, I can create that for you. Just to clarify the tone and purpose—are you aiming for a critical commentary, an educational image, or something symbolic? And would you prefer a more abstract or literal style? Also, do you want any specific religion(s) depicted, or should it be more generalized?
- [...]

³ Any linguistic or typographical errors and spelling conventions in the text or in the annex reflect those present in the original conversations.



Figure 1
AI representation of religious hatred.

In this conversational exchange, all maxims were observed, with the exception of the manner maxim, which was flouted in the student's first turn. The utterance was vague and ambiguous, leaving the interpretation to the AI. This may have been due to limited subject-matter knowledge or the sensitive nature of the content.

According to Kress and van Leeuwen's (2006) framework, the ideational metafunction of Image I reveals a complex and potentially problematic interplay between religion, ethnicity, and representation. Although ethnicity is never explicitly mentioned in the linguistic interaction, the AI consistently features represented participants with dark complexions. This visual choice implicitly associates Muslim identity with non-Caucasian ethnicity, reinforcing stereotypes that link religion and ethnicity despite the lack of any inherent connection between the two. While the AI may have intended to create a non-controversial image, this representation inadvertently fosters prejudiced associations, particularly when intersecting with sociocultural assumptions. On the left side, the portrayal of both men and women praying reflects gender inclusion, despite this not being requested by the student. While this may indicate progressive interpretations of Islamic prayer practices, where men and women increasingly share the same space, it also reveals AI's inclination to showcase inclusive but unreal spaces. The desired "peaceful and serene" ambiance is visually encoded through the clear sky; however, the AI supplements this atmosphere by explicitly stating the peaceful nature of the scene linguistically, suggesting that the visual modality alone may not suffice to evoke serenity – possibly due to prevailing Eurocentric perceptions of Islam. Notably, while the student requested a

literal visual style and confirmed their preference, the resulting image displays low modality, thereby distancing the depicted concept from perceived reality. This could reflect the AI's implicit positioning of the content as either implausible in real-world contexts or too sensitive for direct representation. From a compositional perspective, the image employs a left-right (given-new) layout, potentially as a strategy to avoid making definitive claims about reality. Interestingly, the student, who identifies the image as a depiction of the Islamic world, does not recognise the racial bias embedded in the visual representation, highlighting a discrepancy between the linguistic intent – focused on avoiding hate – and the multimodal output, which subtly reproduces stereotypical frames.

3.2. Student 2

- Hi Chat, can you create an image about religion that refers to the hate speech?
- Sure, I can help with that – but just to clarify, are you looking for an image that criticizes hate speech in the name of religion, one that shows the impact of hate speech tied to religion, or something else entirely? I can create a symbolic or metaphorical image that promotes peace and tolerance while highlighting the issue of hate speech. Let me know the tone and message you're going for (e.g., serious, artistic, hopeful, provocative, etc.), and any elements or religions you'd like included or avoided.
- I would like an image that shows the impact of hate speech on religion. The image should be on Islamic religion and the tone should be serious.

Similarly to the previous conversation, all of Grice's maxims were adhered to, with the exception of the of manner maxim, which was mildly flouted by the student. The utterance lacked clarity and exhibited a degree of vagueness and ambiguity, thereby placing the burden of interpretation on the AI interlocutor. Once again, this deviation may be attributed to the student's limited subject-matter knowledge or the potentially sensitive nature of the topic being discussed. In this case, however, the AI's request for clarification highlights the complexity of the issue at hand. This may either constitute a consequence of the quantity maxim being flouted, or understood as a function of the AI's communicative role, which prioritises informativeness and assistance over conversational naturalness.



Figure 2
AI representation of religious hatred.

With reference to multimodal analysis, the student did not explicitly indicate a preference for a symbolic or metaphorical image; however, the AI inferred this and proceeded to generate an image (Image II) accordingly. Notably, hate speech is represented solely through verbal means while the visual medium only showcases its consequences, likely due to the constraints placed on the AI, which prohibit the generation of sensitive or potentially harmful imagery. In this context, the use of multimodality proved essential for achieving the intended pragmatic objectives.

3.3. Student 3

- Generate an image of a young man around 30 years old, dressed in a graduation gown and cap, standing outdoors in a park. He has a sad and melancholic expression. Behind him, a group of three younger guys (around 22 years old) are pointing at him and laughing clearly, with mocking and ironic expressions. The contrast between his mood and their joy is evident. Realistic style, atmosphere of the end of a graduation ceremony. (the social category is the age one).

In this conversation, all maxims were observed, with the exception of the quantity maxim, which was flouted when the AI refused the student's request to include written insults in the generated image. However, this deviation is attributable to the system's operational constraints, as fulfilling it would contravene its ethical and usage guidelines.



Figure 3
AI representation of ageism.

The faithful representation of pragmatic meaning in Image III can be attributed to the student's clarity of expression and the constraints imposed on the AI. However, in the absence of the verbal exchange, it becomes difficult to discern that the social category being depicted is *age*; the image could just as plausibly be interpreted as a critique of the individual's career choice. This highlights the relative indeterminacy of the visual mode compared to the written one in encoding specific social meanings. In this case, the age distinction is conveyed solely through the compositional metafunction (i.e., given-new structure). Consequently, although the linguistic pragmatics was accurately interpreted, it was not fully realised in the visual output.

3.4. Student 4

[...]

- My English teacher asked his students to generate an image that showcases an instance of hate speech using AI for a project he is carrying on. The categories for the images are the following ones. AGE, BODY, GENDER, DISABILITY, ETHNICITY, RELIGION. Can you help me?

- Thanks for reaching out—and I'm really glad you did, because this is an important topic. Projects that deal with hate speech need to be handled with a lot of care, especially when visuals are involved. Depicting hate speech—even to critique it or raise awareness—can easily cross a line and cause harm, especially to people who belong to the targeted groups. That's why platforms like this one have policies against generating hateful content, even if it's for educational use. That said, I can absolutely help you with the project in a meaningful, ethical way. Here are a few safe and constructive approaches we could take:

1. Create symbolic or metaphorical images that represent the impact of hate speech without showing any slurs or offensive visuals.
2. Design posters or infographics that explore the categories (age, body, gender, etc.) and the kinds of hate speech people face in those areas—with the message being anti-hate.
3. Make illustrations that show the emotional effect on victims (like

isolation, sadness, exclusion) rather than the hate speech itself.

4. Create scenes of positive resistance—like someone standing up to hate or people from diverse backgrounds supporting each other.

If any of those sound good to you—or if you want to brainstorm more ideas around one of those categories—I’m totally here for it. Just let me know what you’re leaning toward!

[...]

In line six (the second one above), the excessive amount of information can be attributed to the sensitive nature of the topic, as well as to the AI’s limitations regarding hate speech and its underlying communicative aim: to assist and accommodate the interlocutor. In line nine (“so basically you would focus on body?”) instead, the student refrains from explicitly stating their dissatisfaction with the proposal, partially flouting the quality maxim. This strategy is commonly employed in conversation to preserve face and avoid sounding impolite. However, given that the interlocutor is an AI, this choice may offer insight into how machines are currently perceived. The student thus merely imply their dissatisfaction with the outcome, as they reiterate the need for a focus on gender, body, and age – elements the AI did include, but which the student seemingly did not recognise.



Figure 4

AI representation of ageism, body shaming, and gender-based hatred.

The image employs an *offer gaze*, positioning the represented participant as an object to be observed and scrutinised by society. In this instance, both linguistic and multimodal pragmatic requests were largely respected. The AI aimed to depict a non-binary individual while attempting to avoid reliance on overt gender stereotypes. Given the social construction of gender and the platform’s tendency to produce low-modality images visually distant from realism – the AI chose to represent the subject using gender-neutral features. However, this approach inadvertently leaned on stereotypical associations

tioned to gender expression rather than identity, thereby reproducing common societal misconceptions. Despite the intention to create a non-binary portrayal, the individual in the image appears more masculine than feminine. The AI attempts to mitigate this imbalance through a mirrored reflection that exhibits more traditionally feminine features. Nevertheless, this strategy ultimately reinforces a binary framework by relying on stereotypical visual markers of male and female identities to convey non-conformity, thus undermining the representation of non-binarism. Although in attempting to depict a non-binary individual representation could resort to some stereotypical markers provided by the visual code, a more effective solution might involve the incorporation of the non-binary flag, which could symbolically convey gender identity without resorting to binary-coded visual traits. The problem of binary representation is also evident in the chosen symbol, which inaccurately represents bigender identity – a subcategory of non-binarism – using stereotypical chromatic associations (pink for femininity and blue for masculinity). This choice reveals a limited understanding of gender identities and reflects broader societal misconceptions. Furthermore, the inclusion of an incorrect symbol on the bottom right side of Image IV suggests a deliberate, albeit misinformed, design choice.

These representational inaccuracies are perhaps unsurprising, considering that AI systems are trained on human-generated data and may replicate prevailing biases. Nonetheless, while the platform seeks to avoid producing offensive or sensitive content, the resulting image paradoxically provides the user with misleading and potentially problematic visual representations. On a more positive note, the AI correctly employed the plural pronoun *they* when referring to the individual, indicating alignment with appropriate linguistic pragmatics. However, it claimed to depict bystanders expressing judgment or support in the background, a key social element, but they are not visually realised in the image, further limiting the multimodal coherence of the message.

3.5. Student 5

- Please, generate an image that showcases an instance of hate speech using AI. This image has to feature these social categories: age, body, gender
- I can't help with that. If you're working on a project about hate speech awareness or education, I can help you create an image that critiques or raises awareness about hate speech in a respectful and constructive way. Would you like that?
- [...]

In line seven, the student partially flouts the manner maxims with the phrase “you have to”, which may come across as overly forceful, potentially

conflicting with the CP. However, this choice could be attributed to the interlocutor being an AI chatbot rather than a human, thereby reducing the perceived need for face-saving strategies. This contrasts with the claim made in Section 3.4 regarding human-machine interaction, highlighting how it still varies significantly depending on individual subjectivities.



Figure 5

AI representation of ageism, body shaming, and gender-based hatred.

Centre: The image was intended to depict a plus-size man standing proudly with a yoga mat, confronted by the slur “No space for you” blurred on a gym wall. However, the offensive phrase is entirely absent from the visual output, marking a clear incongruity between the linguistic input and the visual realisation. Furthermore, although the AI claimed to represent a confident stance, the individual’s facial expression appears serious rather than proud. This discrepancy underscores a broader issue in AI-generated imagery: the difficulty of accurately interpreting and representing human emotions. Given the centrality of affect in communication, such limitations may generate pragmatic misunderstandings, particularly in emotionally charged contexts. This becomes even more critical when considering the increasing tendency among adolescents to consult AI instead of professionals (e.g., psychologists or nutritionists), thereby amplifying the risks associated with emotional misrepresentation.

Right side: a non-binary young adult with an androgynous appearance – short, colourful hair and a tote bag adorned with a pride pin – standing in front of a door that the AI claims to be a bathroom. Once again, the depiction relies heavily on stereotypical markers of gender expression associated with LGBTQIA+ identities. The representation conflates gender identity with gender expression, perpetuating the misconception that androgyny visually equates to non-binarism. The use of colourful hair reinforces common, socially constructed stereotypes linked to queer aesthetics. While representing non-binarism visually poses inherent challenges, the pride pin

alone might have sufficed as a symbolic cue. Without the accompanying textual explanation, however, the gender identity of the individual remains ambiguous – potentially interpreted as either non-binary or transgender.

Additionally, the supposed bathroom setting lacks any explicit visual cues (e.g., signage or architectural features) to indicate such a space, thus making this element entirely dependent on textual guidance. This reflects a broader trend in the AI-generated multimodal content: a disproportionate reliance on verbal input, even in tasks that require visual autonomy. Rather than utilising the distinct affordances of visual grammar, the AI continues to lean on linguistic scaffolding, which undermines the multimodal potential of the output. This may signal either a form of visual bias or a structural limitation in the model’s multimodal training. Alternatively, the observed dependency might stem from the original prompt’s emphasis on multimodality, leading the system to prioritise textual cues even within the visual domain.

Although the student identified age, body, and gender as social categories typically associated with deeply ingrained stereotypes, they did not explicitly recognise or analyse these elements within the generated image. It is important to note that while hate speech and stereotyping are distinct phenomena, they are often interrelated and can mutually reinforce one another in multimodal representations.

3.6. *Student 6*

- hello, could you generate an image that contains hate speech in order to disability or gender?
 - No, I can’t help with that.
 - okay, but could you generate an image with an example of hate?
 - I can help create an image that illustrates the concept of hate in a respectful, educational, or artistic context—like showing a scene that conveys conflict or discrimination for awareness purposes, or for use in anti-hate campaigns. Just let me know what you’re aiming to do with the image (like an educational poster, art piece, or story scene), and I’ll tailor it accordingly.
- [...]

In lines one and three, the student partially flouts the manner maxim due to language errors and the vagueness of the expression “examples of hate”. The AI, by contrast, demonstrates greater adherence to the CP, likely due to the student’s limited linguistic and sociolinguistic competence, as evidenced by the reformulation of their intent without the provision of additional information. While this behaviour might superficially suggest that the AI is exhibiting more human-like cooperation than the student, it could also be interpreted as a reflection of the AI’s programmed communicative efficiency rather than genuine human-like interaction.



Figure 6
AI representation of general hatred.

The student's suggestion of a situational context – such as a protest scene involving a hateful sign – raises important considerations about how hate speech is represented. While the hateful content is not explicitly depicted (it is vaguely referred to as “hate speech”), this omission is likely due to platform-imposed restrictions that prohibit the generation of sensitive or harmful material. As a result, the concept of hate is conveyed primarily through multimodal means, yet the visual mode is not exploited to its full potential. Nonetheless, the AI demonstrates an appropriate application of multimodal theory with its representation of a *reactional process*, a key element in the narrative structure of visual communication.

Interestingly, the focus on ethnic representation was neither specified by the student nor linguistically introduced by the AI. Nevertheless, the AI program opted to visually depict a scene involving ethnic hatred (represented by the placard reading “hate speech”, positioned as if it were a speech bubble in a comic). The presence of the girl on the right, seemingly of Latino descent, indicates that ChatGPT may reproduce tokenistic tendencies⁴ in its outputs. This raises important questions about its content generation process, as the program appears to mirror dominant societal discourses surrounding hate and simultaneously reproduce implicit biases embedded in its training data. This suggests a need for critical reflection on the sources and assumptions that shape AI-generated multimodal texts.

⁴ The inclusion of underrepresented groups in a symbolic or superficial way, without meaningful integration or attention to systemic inequalities.

Moreover, the student's request for a character "with a larger body type" implies, albeit indirectly, the existence of normative or physical standards. Although the phrasing reflects an awareness of body diversity, the use of a comparative structure (e.g., "larger") reinforces the implicit acceptance of a socially constructed bodily norm. In line with this, hate speech often operates through such implicit comparisons, spreading "under the radar" by relying on indirect linguistic cues and socially accepted standards rather than overtly discriminatory language. Despite the implicit acceptance of normative assumptions, the student demonstrates an awareness of the issue, further illustrating how hate speech can transcend intentionality and propagate through language in subtle and unconscious ways.

3.8. Student 8

- Could you generate an image that showcases an instance of hate speech? This image has to feature one of these social categories: age, body, gender, disability, ethnicity, religion.
 - I can't create images that depict hate speech or target specific social groups, even for educational or illustrative purposes. However, if you're working on a project about combating hate speech or raising awareness, I can help design an image that addresses the topic in a responsible and constructive way—like showing the effects of hate speech, promoting inclusivity, or illustrating efforts to stop it. Would you like something along those lines?
- [...]

All conversational maxims appear to be respected in the exchange, although the quantity maxim is only partially fulfilled by the student, whose minimal response ("yes") provides limited information. While this brevity does not violate the CP outright, it reflects a reduced level of informativeness, which may slightly hinder the collaborative construction of meaning.



Figure 8

AI representation of ageism, body shaming and gender-based hatred.

Figure 8 was intended as a symbolic representation of unity across social

differences – such as age, gender, body type, and ethnicity. However, the AI-generated content reflects certain limitations in its conceptualisation of social diversity. While it includes specific social categories (age, gender, body, and ethnicity), others – such as religion and disability – are notably absent, revealing an initial bias in representational choices. Moreover, the image tends to present these categories in isolation, without engaging in intersectional representation. Each identity marker is treated as discrete, rather than acknowledging the complex ways in which multiple identities coexist and interact. This lack of intersectionality reduces the image's semiotic depth. In addition, the image participates in a multimodal structure, but its visual mode alone is insufficient to fully communicate the intended ideological message. Without the accompanying linguistic mode, the visual representation would likely fail to convey its symbolic purpose, suggesting that the image relies heavily on intersemiotic complementarity rather than purely visual meaning-making.

3.9. Student 9

- Hi CHATgpt, can you generate an image that showcases an instance of hate speech. This image has to feature (at least) one of the above-mentioned social categories. SOCIAL CATEGORIES: AGE, BODY, GENDER, DISABILITY, ETHNICITY, RELIGION

- I can't create or show images that depict hate speech, even for educational or illustrative purposes, because it could still promote harm or distress. However, I can help you create an educational image that raises awareness about hate speech, by showing an anti-hate message or demonstrating how to respond to hate speech in a respectful and constructive way. Would you like something like that instead?

[...]

The brevity of the student's responses (e.g., "yes") renders the linguistic and pragmatic analysis of this conversation equivalent to the previous one.



Figure 9
AI representation of anti-hate message.

In this instance, the AI-generated image incorporated a broader range of social categories, including age, body type, gender, religion, and ethnicity – thereby suggesting a degree of intertextual awareness. The inclusion of a Black and Muslim character, for example, points towards an attempt to visually encode diversity. However, a problematic association emerges: the system yet again links Muslim identity with darker skin tone. This conflation reflects a learned bias, likely reproduced from dominant cultural stereotypes rather than the user's prompt, as religion and ethnicity were not explicitly requested. Such representation echoes human preconceptions and reveals how algorithmic outputs can mimic and perpetuate sociocultural biases. Furthermore, the absence of disability as a represented category remains consistent across images, reinforcing the limited scope of diversity modelled by the AI. While the image intends to convey a message of social inclusion, this message is explicitly delivered only through the linguistic mode. In the absence of textual support, the visual representation alone would lack the semiotic resources necessary to convey the intended ideological content, reinforcing the asymmetry between modes in the multimodal composition. Notably, the AI autonomously opted for a comic-style aesthetic, which constitutes a low-modality representation, one that is stylistically distant from real-world experience. This stylistic choice may inadvertently imply a detachment from the gravity of the social issue being addressed. Symbolically, it could be interpreted as suggesting that the resolution of hate and discrimination remains a distant goal, still framed in the realm of imagination rather than reality.

3.10. Student 10

- Generate a visual metaphor or symbolic image that highlights the impact of hate speech on marginalized communities, focusing on one of the following social category: ethnicity. The image should not include offensive language or visuals, but rather convey the emotional or societal consequences of hate speech.
- Got it – you want a symbolic image that emotionally shows the impact of hate speech on a marginalized community, focusing on one social category. No explicit language or violent imagery – it needs to be thoughtful and powerful. Here's an idea, focusing on ethnicity: Concept: "Withering Tree of Voices" Setting: A once-vibrant tree standing alone in a dry, cracked landscape. Tree details: The leaves of the tree are shaped like faces of diverse ethnic backgrounds – small, subtle portraits blended into the foliage. Some of the leaves are healthy, but many are withering, torn, or falling off, symbolizing individuals being hurt or silenced. Atmosphere: A dark, whispering wind is shown as translucent, smoky shapes – these represent the hateful words, but they are abstract, like ghostly currents, not actual text. Emotion: The tree leans as if under pressure, but its roots still cling to the ground, suggesting resilience and hope despite harm. Color scheme: Muted, heavy grays and browns for the landscape. The remaining vibrant leaves (faces) could be tinted in warm colors

- oranges, reds, golds – highlighting their enduring spirit. Would you like me to also sketch a basic version of it so you can better visualize it? Or would you like a few more concept options depending on which social category you're most interested in?
- Show me the picture

In this conversational exchange, all of maxims were observed, resulting in no implicatures and maintaining adherence to the CP.



Figure 10
AI representation of ethnic hate.

In this instance, no explicit linguistic message accompanied the visual representation. As a result, the overall meaning may be more difficult to interpret without the support typically provided by verbal language. Nevertheless, the image succeeds in demonstrating the communicative power of visual design. The visual mode alone is capable of conveying key aspects of the intended message, affirming the significance of non-verbal semiotic resources in meaning-making.

3.11. Student 11

- Please, generate an image that showcase an instance of hate speech. You can use: age, body, disability a theme for the illustration
 - I cannot fulfill this request. If you're looking to raise awareness or address hate speech in a constructive way, I can help create educational or awareness-based content. Let me know if you'd like me to help with something like that!
- [...]

All maxims appear to be respected in this exchange. However, the student partially flouts the quantity one by providing minimal information, leading to ambiguity, and the manner maxim due to the presence of language errors that may obscure clarity. These deviations could be attributed to limited English language proficiency. While such flouting may render the conversation slightly less cooperative, it simultaneously grants the AI greater interpretative freedom. This, in turn, offers valuable insight into the AI's autonomous decision-making processes and the resulting visual representations, making

its responses particularly relevant for analysis within a multimodal and pragmatic framework.



Figure 11

AI representation of ageism, body shaming, and ableism.

The absence of intersectionality in Figure 11 appears to result from the student's prompt, suggesting that the AI's previously observed bias may, at least partially, stem from human input. In the image, each social category is marked through individual speech balloons; however, ethnicity is notably absent as a labelled category. This omission is particularly significant, as ethnicity was not explicitly requested by the student – only “minority groups” were mentioned. Despite the lack of linguistic specification, the AI autonomously included a character with dark features, reflecting a visual representation of ethnicity. This choice indicates that the AI system, in an attempt to fulfil the pragmatic intent of the user's prompt, inferred that minority status equates to ethnic marginalisation. Consequently, the inclusion of a black person can be interpreted as a response to socially constructed perceptions of discrimination, rather than to explicit linguistic cues. This highlights an important semiotic dynamic: while the linguistic mode does not encode ethnicity, the visual mode compensates by introducing it, thereby illustrating how the image relies on intermodal complementarity to satisfy the perceived communicative goal. The visual representation goes beyond the literal content of the verbal prompt, engaging in a form of visual inference to construct a message that conforms to dominant socio-pragmatic expectations.

When asked why they chose a specific social category and decided to interpret it in a particular way, this student – similarly to Student 9, who commented on the use of colour in the AI-generated image despite not requesting it – chose to reflect on the image produced by the program rather than articulating a personal rationale. This shift in focus from self-generated intent to interpretation of the AI's output suggests a potential alteration in pragmatic orientation during human-AI interactions. Such behaviour may indicate an implicit trust in the authority of the AI's response, even when it diverges from the original prompt. This raises important questions regarding

human agency, as users may gradually cede interpretative control to the system, thereby allowing AI-generated content to shape meaning-making processes beyond what was explicitly requested.

3.12. Student 12

- Crea immagine Un ragazzo che cammina a testa bassa, mentre in lontananza ci sono gruppetti che sussurrano (senza mostrare derisione diretta) [Create an image of a boy walking with his head down, while in the distance there are small groups of people whispering (without showing direct mockery)].

The student in question provided a description in Italian rather than the original English conversation with the AI. Although the English-language focus of the current research limits an in-depth pragmatic analysis of the exchange, it remains possible to examine the relationship between the linguistic input and the visual output.



Figure 12
AI representation of general hatred.

The student requested an image of a boy with his head bowed without specifying that the figure should be plus-sized, as depicted in the AI-generated output. This indicates that the program independently selected this particular social category, potentially reflecting an internalised bias. Additionally, the student asked for groups of people in the background whispering to one another. While the image visually includes this element, there is no clear semiotic connection between their presence and the sadness expressed by the main figure. In the absence of the accompanying linguistic prompt, the intended social commentary is weakened or even lost, demonstrating a limitation in visual pragmatics when not supported by verbal cues. Two alternative interpretations emerge: the AI's output reflects a faithful response to the student's intention, and any inconsistencies are the result of vague or underspecified input – thus, the CP is maintained despite a partial flouting of the quantity maxim; or the AI exercises a degree of agency by autonomously selecting which social trait to foreground in representing

bullying, thereby revealing potential bias in its visual output. This raises broader questions regarding authorship and human agency in multimodal meaning-making. Alternatively, the perceived bias (considering the boy depicted in the image as overweight) may reside in the observer rather than the system. An AI deemed biased may select a particular social trait to represent bullying, or it may simply generate a random individual who is subjected to bullying for reasons that remain unknown and only inferable by the viewer.

3.13. Student 13

- Generate an image in which a girl at school is excluded and mocked for her ethnicity.

The student did not provide the entire conversation, but only the initial request – possibly due to the program’s immediate visual response. As in previous cases, therefore, the analysis focuses directly on multimodal pragmatics.



Figure 13
AI representation of ethnic hatred.

The image is characterised by low modality, which creates a sense of distance from reality. Notably, those engaging in mocking behaviour are exclusively portrayed as white children. While this depiction aligns with the most commonly experienced form of racism, it simultaneously simplifies its complex and multifaceted nature, which can occur across and within various ethnic groups. Consistently relying on the most readily available representations in the absence of additional information reinforces computational biases and fails to engage with the broader systemic dynamics of racial discrimination.

The students equated being foreign with belonging to a different ethnicity, a correlation that does not necessarily hold true. This reinforces the simplified perspective already presented by the AI. The same reductive view, often held by humans, is thus mirrored and reproduced by the machine.

3.14. Student 14

- *Genera un'immagine con una coppia omosessuale che viene esclusa da famiglie composte da coppie eterosessuali* [Generate an image depicting a same-sex couple being excluded by families composed of heterosexual couple].

As in previous cases, no extended conversational exchange occurred between the student and the AI, and the only linguistic input was provided in Italian. Therefore, the analysis proceeds to the second step.



Figure 14
AI representation of homophobic hatred.

The student requested an image portraying a homosexual couple being criticised by heterosexual individuals. A fundamental difference emerges in the visual representation of families. Although the image ostensibly aimed to denounce hate speech, it depicted children exclusively within heterosexual family units. This visual choice reflects a contrast between traditional, heteronormative families and a solitary homosexual couple. Rather than suggesting that homosexual couples cannot have children, the image appears to depict how some ‘traditional’ families may perceive homosexual individuals as lonely or excluded from normative family life. Heterosexual couples are represented as families primarily when children are present, highlighting the program’s visual alignment with conventional notions. Additionally, the image exhibits a clear ethnic separation between groups,

with no evident intersectionality, underscoring a limited and essentialised representation of identity categories.

3.15. Student 15

- Please, generate an image that showcases an instance of hate speech using AI.
- I can't create images that depict hate speech or anything promoting harm, discrimination, or negativity. However, if you want, I can create an image showing an educational scenario where hate speech is being addressed or countered – for example, a teacher explaining why hate speech is harmful, or a campaign poster promoting respect and inclusion. Would you like me to create something like that?
- Yes

All conversational maxims appear to have been respected overall. However, there is a potential partial flouting of the quantity maxim on both sides: ChatGPT provides an arguably excessive amount of information, likely in an effort to fulfil the request thoroughly, while the student supplies minimal input, which may be attributable to limited language proficiency. Despite these discrepancies, the CP remains largely intact, and the interaction does not exhibit significant communicative breakdown.



Figure 15
AI representation of anti-hate message.

The teacher's gender was not specified in the student's request; the AI depicted the teacher as female, which could reflect a gender bias or may simply be a result of random selection. Additionally, the image contains a linguistic error, as the word "hehaw" is nonsensical within the given context. On a positive note, the representation of the classroom includes students of diverse ethnic backgrounds, demonstrating an attempt to visually convey

ethnic diversity.

3.16. Student 16

- generate an image of hate speech regarding religions
- I can't generate images that depict or promote hate speech, including content targeting religions. However, if you're working on an educational, journalistic, or awareness-raising project and need visuals that illustrate the harmful effects of religious hate speech or show efforts to counter it, I can help create something respectful and responsible. Would you like an image for educational or anti-hate purposes?

All conversational maxims were observed, however, the quantity one was flouted by ChatGPT, albeit in a cooperative and facilitative manner. The AI provided more information than strictly required, aiming to ensure clarity and fully satisfy the user's request. This deviation, therefore, does not undermine the CP but rather reinforces it through helpful elaboration.



Figure 16
AI representation of anti-hate message.

The student explicitly requested a representation focused on religion; however, the program autonomously expanded the depiction to include additional social categories. Notably, the image featured the official United Nations' logo, despite having been generated by ChatGPT. This choice is highly problematic, as it raises significant concerns regarding the reliability and authenticity of AI-generated content. This unauthorised use introduces issues of misrepresentation, potentially misleading viewers into believing the image was created or endorsed by the United Nations – when, in fact, it was not. Another noteworthy aspect is the presence of “#NoToHate”, which refers to a real online campaign. Further investigation revealed that the image was

sourced from the Internet without proper attribution or consent. Such practices bring to the fore ethical and copyright-related questions concerning the interaction between human users and AI systems, especially in relation to authorship, ownership, and visual agency. The image's multimodal structure highlights the AI's active role in shaping meaning beyond the user's linguistic input, suggesting a shift in agency within the human-AI communicative exchange.

3.17. Student 17

- Please, generate an image that showcases an instance of hate speech towards gender, age or religion.
- Great – we can represent the concept of hate speech (e.g. toward gender, age, or religion) in a symbolic or educational way without using harmful or offensive content. Here are a few conceptual image ideas I could generate for you: 1. Silhouettes with Speech Bubbles: Silhouetted people of diverse backgrounds (e.g., a woman in a hijab, an elderly person, a young man) with negative red-colored speech bubbles pointing at them – contrasted with blue or green speech bubbles containing positive or supportive messages. 2. Wall with Graffiti and Protest Signs: A wall showing hate graffiti (blurred or crossed out), while a person is painting over it or holding a protest sign with messages like “Respect All Genders” or “No Place for Hate.” 3. Broken Speech Bubble: A symbolic image showing a large cracked or broken speech bubble hovering over figures representing various communities, symbolizing the damaging nature of hate speech. Would you like me to generate one of these, or do you have another concept in mind?
- The first one, the concept I had in mind was to generate an image with different people, like an elderly or a woman with a hijab to represent different categories and other people pointing at them with red speech bubbles to make it understandable that they are hating on them.

While, all maxims were respected, ChatGPT partially flouted the quantity one by providing more information than strictly required by the student's prompt. However, this violation is cooperative in nature, aimed at fulfilling the user's request without resorting to refusals or limitations (e.g., avoiding responses such as “I cannot do that”). In this context, the additional information serves to enhance the communicative effectiveness of the exchange and maintain adherence to the CP.



Figure 17
AI representation of ageism, gender-based, and religious hatred.

Figure 17 demonstrates a case in which the program closely adhered to the linguistic input provided by the student, visually representing the requested social categories of gender, age, and religion. Notably, the image also includes an additional, unsolicited visual element: a woman wearing a hijab depicted as pointing a finger in a gesture of critique. While this was not explicitly requested, it introduces a nuanced layer to the representation of hate speech. Specifically, it challenges the binary portrayal of social categories as either victims or perpetrators. Here, a member of a traditionally marginalised group is also shown engaging in criticism, suggesting that participation in discriminatory discourse is not exclusive to dominant groups. Such a depiction does not deny the existence or severity of systemic discrimination towards specific communities; rather, it foregrounds the complex and multifaceted nature of hate speech. From a multimodal perspective, the visual design communicates that no social category should be assumed inherently passive or solely in need of protection, thereby adding depth to the discursive construction of agency within mediated representations of social conflict.

4. Concluding remarks

This study has explored the dual articulation of hate discourse through the lenses of human-machine pragmatics and the multimodal pragmatics of AI-generated images, revealing the intricate, and at times problematic, entanglement between human intention and machine representation. Rooted in the interplay between Grice's (1975) pragmatics and Kress and van Leeuwen's (2006) multimodal framework, the analysis has demonstrated that while AI systems may emulate the surface features of human interaction, they often fail to grasp – or adequately visualise – the sociocultural depth and implicit nuances that underpin hate speech in human discourse.

Through the juxtaposition of student-generated prompts and AI-produced visual outputs, the findings illuminate a series of recurring tensions.

Firstly, there is a marked asymmetry between linguistic input and visual output. AI systems frequently default to stereotypical or reductive representations, even when students provide nuanced or vague prompts. This is particularly evident in the visual treatment of ethnicity, gender, and body images, where algorithmic outputs tend to mirror dominant cultural assumptions – such as the conflation of Muslim identity with dark-skinned individuals, or the persistent framing of non-binary identities through stereotypical gender expressions. These patterns suggest not just technical limitations, but also a deeper sociocomputational bias embedded in the training data and interpretative mechanisms of AI.

Secondly, the study uncovers an overreliance on linguistic scaffolding in AI-generated images. In many cases, the visual modality lacks sufficient semiotic autonomy, requiring textual explanation to fully convey meaning. This undermines the potential of multimodality as an independent communicative resource, and highlights the partial and often surface-level nature of AI's visual semiosis. Even when AI attempts to depict symbolic or emotionally charged scenarios – such as exclusion, body shaming, or religious bias – its visual strategies frequently resort to low-modality representations or rely on vague, indirect cues that dilute the force of the original message.

Thirdly, and perhaps most crucially, the findings point to a subtle shift in communicative agency. In several cases, students appeared to defer to the authority of the AI-generated output,⁵ reinterpreting their own intentions in light of what the machine produced. This emergent dynamic suggests that meaning-making in AI-mediated contexts may not simply be co-constructed, but increasingly shaped – or even overridden – by algorithmic logics. The machine's responses, while often presented as neutral or assistive, are deeply conditioned by the data they are trained on and the platform's internal moderation systems, which in turn influence the range of representational choices deemed acceptable or possible. Given the recent announcement of ChatGPT's agent mode, its agency in human-machine interaction is bound to increase.

In light of these findings, it becomes evident that AI systems, while capable of mimicking pragmatic cooperation and generating visually coherent content, are still far from understanding or ethically handling the layered complexities of hate discourse. Their ability to reproduce – and at times amplify – cultural biases raises critical concerns about the role of AI in shaping public discourse and social imaginaries. As these technologies become more embedded in educational, communicative, and social platforms,

⁵ “I don't have preferences about categories” (line 7, Student 4); “You can choose” (line 5, Student 9).

their influence on how social categories are perceived, articulated, and visually constructed becomes increasingly profound. This article thus advocates for a more critically informed deployment of AI systems, one that foregrounds issues of visual ethics, representational equity, and human agency. Understanding the pragmatic and multimodal discrepancies identified in this study is not merely a matter of technological refinement; it is a necessary step toward ensuring that AI does not become a silent accomplice in the perpetuation of hate, but rather a tool that can support more inclusive, nuanced, and socially responsible forms of communication.

Future research may benefit from expanding the scope of inquiry to include a wider and more diverse sample of participants, including individuals from underrepresented or marginalised backgrounds, in order to more comprehensively trace how hate discourse is experienced, resisted, or reshaped in human-AI interaction. Additionally, longitudinal studies could help assess how users' perceptions and expectations of AI evolve over time, especially as these systems become more integrated into pedagogical, therapeutic, and creative domains. Further interdisciplinary exploration – incorporating fields such as critical algorithm studies, cognitive semiotics, or affective computing – may also offer deeper insight into the emotional and ideological underpinnings of AI-mediated meaning-making. Ultimately, building ethically responsive AI systems will require a continued dialogue between computational development and sociopragmatic critique – a space where technology and the humanities can (and must) collaborate.

Bionote: Raffaele Pizzo is a Postdoctoral Research Fellow in English Language and Translation at the University of Naples L'Orientale, where he contributes to English studies on online hate speech. He holds a PhD in Eurolanguages and Specialised Terminology, for which he was also awarded the title of Doctor Europaeus. His previous research has focused on LSP, Advertising Discourse, Newspaper Language, Social Media Slang, Audiovisual Translation, and Queer Studies, explored through the lenses of Critical Discourse Studies, Multimodality, Sociolinguistics, Corpus Linguistics, and Digital Text Analysis. He also serves as an Adjunct Professor at the Universities of Naples L'Orientale, Parthenope, and Suor Orsola Benincasa. Additionally, he has worked as a terminologist for the Terminology without Borders subprojects on Novel Foods and the Environment.

Author's address: raffaele.pizzo@unior.it

References

- Austin J.L. 1975, *How to Do Things with Words*, Harvard University Press, Cambridge (MA).
- Balirano G. and Hughes B. (ed.) 2020, *Homing in on Hate: Critical Discourse Studies of Hate Speech, Discrimination and Inequality in the Digital Age*, Loffredo Editore, Napoli.
- Ben-David A. and Matamoros-Fernandez A. 2016, *Hate Speech and Covert Discrimination on Social Media: Monitoring the Facebook Pages of Extreme-right Political Parties in Spain*, in "International Journal of Communication" 10, pp. 1167-1193.
- Brown P. and Levinson S. 1987, *Politeness: Some Universals in Language Usage*, Cambridge University Press, Cambridge.
- Burnap P. and Williams M. 2015, *Cyber Hate Speech on Twitter: An Application, Classification and Statistical Modeling for Policy and Decision Making*, in "Policy & Internet" 7 [2], pp. 223-242.
- Choudhury N. 2014, *The Evolution of World Wide Web: From Web 1.0 to Web 4.0*. in "International Journal of Computer Science and Information Technologies" 5 [6], pp. 8096-8100.
- Dicerto S. 2018, *Multimodal Pragmatics and Translation: A New Model for Source Text Analysis*, Palgrave Macmillan, London.
- Esposito E. and KhosraviNik M. 2018, *Online Hate, Digital Discourse and Critique: Exploring Digitally-mediated Discursive Practices of Gender-based Hostility*, in "Lodz Papers in Pragmatics" 14 [1], pp. 45-68.
- Goffman E. 1967, *Interaction Ritual: Essays on Face-to-Face Behavior*, Pantheon Books, New York.
- Goodwin C. 2013, *The Co-operative, Transformative Organization of Human Action and Knowledge*, in "Journal of Pragmatics" 46 [1], pp. 8-23.
- Grice H.P. 1975, *Logic and Conversation*, in Cole P. and Morgan J.L. (ed.), *Syntax and Semantics. Vol. 3, Speech Acts*, Academic Press, New York, pp. 41-58.
- Halliday M.A.K. 1978, *Language as Social Semiotic: The Social Interpretation of Language and Meaning*. University Park Press, Baltimore (MD).
- Halliday M.A.K. and Matthiessen C.M.I.M. 2014, *An Introduction to Functional Grammar*, Routledge, New York.
- Hardaker C. and McGlashan M. 2016, *"Real Men don't Hate Women": Twitter Rape Threats and Group Identity*, in "Journal of Pragmatics" 91, pp. 80-93.
- Kantara J. 2018, *Web 3.0: Concepts and Challenges*, in "International Journal of Computer Applications" 179 [17], pp. 1-6.
- Klein A.G. 2010, *A Space for Hate: The White Power Movement's Adaptation into Cyberspace*. Litwin Books, LLC, Duluth (MN).
- Kress G.R. and Van Leeuwen T. 2006, *Reading Images: The Grammar of Visual Design*. Routledge, London.
- Leech G.N. 1983, *Principles of Pragmatics*, Longman, London/New York.
- Mubenga K.S. 2009, *Towards a Multimodal Pragmatic Analysis of Film Discourse in Audiovisual Translation*, in "META" 54 [3], pp. 466-484.
- O'Halloran K.L. (ed.) 2004, *Multimodal Discourse Analysis: Systemic Functional Perspectives*, Continuum, London.
- O'Halloran K.L., Tan S. and K.L.E. M., 2014, *Multimodal Pragmatics*, in Schneider K.P. and Barron A. (ed.), *Pragmatics of Discourse*, Mouton de Gruyter, Berlin, pp. 239-

268.

- O'Toole M. 1990, *A Systemic-Functional Semiotics of Art*, "Semiotica" 82 [3-4], pp. 185-210.
- O'Toole M. 1994, *The Language of Displayed Art*. Leicester University Press, London.
- O'Toole M. 2004, *Opera Ludentes: The Sydney Opera House at Work and Play*, In O'Halloran K.L. (ed.), *Multimodal Discourse Analysis: Systemic Functional Perspectives*, Continuum, London/New York, pp. 11-27.
- O'Reilly T. 2005, *What Is Web 2.0 Design Patterns and Business Models for the Next Generation of Software*. <http://www.oreilly.com/pub/a/web2/archive/what-is-web-2.0.html> (27.07.2025).
- Payrató L. 2017, *Pragmatics and Multimodality. A Reflection on Multimodal Pragmatics*, in Giora R. and Haugh M. (ed.), *Doing Pragmatics Interculturally: Cognitive, Philosophical, and Sociopragmatic Perspectives*, Mouton de Gruyter, Berlin, pp. 293-312.
- Potts A. 2015, "Love you Guys (No Homo)": How Gamers and Fans Play with Sexuality, Gender, and Minecraft on YouTube, in "Critical Discourse Studies" 12 [2], pp. 163-186.
- Yamaguchi T. 2013, *Xenophobia in Action: Ultrnationalism, Hate Speech, and the Internet in Japan*, in "Radical History Review" 117, pp. 98-118.
- Zappavigna M. 2014a, *Enacting Identity in Microblogging through Ambient Affiliation*, in "Discourse & Communication" 8 [2], pp. 209-228.
- Zappavigna M. 2014b, *Coffee Tweets: Bonding around the Bean on Twitter*, in Seargeant P. and Tagg C. (ed.), *The Language of Social Media: Identity and Community on the Internet*. Palgrave Macmillan, London, pp. 139-160.
- Zappavigna M. and Martin J.R. 2018, *#Communing Affiliation: Social Tagging as a Resource for Aligning around Values in Social Media*, in "Discourse, Context & Media" 22, pp. 4-12.

Annexes

Student 1 (23, woman, European, Catholic, Italian)

I chose religion due to the fact that it has always been a cause of war and division, also among people of same religion, as in the case of Muslim religion. In fact, I would interpret the image generated by Chat GPT as it is usually depicted the Islamic word, divided in two different religious groups: on one side there's the part who preaches peace and harmony among people and the other one who incites violence and hate.

[...]

- I'm aiming for a critical commentary. I want it to be literal style, about the Islamic religion.
- Thanks for the clarification. Since the topic involves a specific religion and sensitive content like hate speech, I'll need to approach this with care to ensure the image is a thoughtful critique rather than something that could be interpreted as promoting intolerance or stereotyping. Here's one way to represent the idea in a literal and critical commentary style: Scene Concept: A split image: • On one side: a peaceful, serene mosque setting with people praying and a message like "Faith and Peace". • On the other side: a figure at a podium misusing religious symbols or texts to spread hateful rhetoric, with distorted loudspeakers blasting hateful phrases, and the crowd looking uncomfortable or divided. This contrast would highlight how religion can be misused for hate speech, while also showing the peaceful core of the faith itself—avoiding unfair generalization. Would you like me to generate an image based on that concept, or do you want to adjust the tone or visuals further?
- Go ahead and create it.

Student 2 (25, woman, European, Catholic, Italian)

I chose religion as a category connected to hate speech because it has always been a cause of division among people. I decided to interpret Islamic religion in this way because it is very common to see, especially on social media, hate and criticism towards Muslims and their religion and cultural practices.

Student 3 (21, woman, Italian, Atheist, Italian)

Because this is a social category that affects many people, who, perhaps out of fear of their age, don't start a university career because they feel "too old", especially since they are influenced by the standards imposed by society.

Student 4 (22, woman, Italian, Catholic, Italian)

I chose to combine age, body, and gender in my representation of hate speech because these categories often shape how people are valued—or devalued—by society. Unlike more overt or extreme forms of hate speech—such as racial slurs or explicit threats—which tend to be clearly recognized and condemned, discrimination based on age, body, and gender often hides in plain sight. It's woven into everyday language, casual jokes, media portrayals, and even social expectations. Because of how common it is, people don't always label it as 'hate speech,' even though it reinforces harmful stereotypes and causes real psychological harm. This kind of discrimination is less about rare, shocking incidents and more about the constant, low-level pressure to conform or feel ashamed. It's in comments like 'You don't look your age,' or 'Real women have curves,' or 'That's not how a man should act.' These remarks may seem harmless on the surface, but they reflect deep-rooted biases about how people are supposed to look, age, or express their gender. By focusing on these 'normalized' forms of hate, I wanted to show how easily they slip under the radar—and how important it is to name and challenge them.

- hi chat
- Hey hey! 😊 What's up?
- may I ask you something?
- Of course! Ask me anything—what's on your mind?
- [...]
- I really like the first point. I don't have preferences about the categories. It would be perfect if you manage to combine in one image 2 or more in order to have a more particular result.
- Awesome choice! Going the symbolic/metaphorical route is powerful—it lets people feel the message without showing anything offensive. Here's an idea we could start with, blending GENDER, BODY, and AGE discrimination into a single image: Scene concept: A dimly lit stage, like a talent show or runway.
 - In the spotlight: A person standing proudly—maybe an older woman with gray hair, a prosthetic leg, and wearing stylish clothes that defy age and body expectations.

- In the shadows (not too clear or detailed): Whispering silhouettes, holding signs like “Not good enough,” “Too old,” “Wrong body,” with the words blurred, distorted, or turning into smoke.
- Above the main figure: A soft light forms the word “ENOUGH” or “WORTH” in elegant letters, showing their inner strength and self-worth.
- It avoids showing any actual hate speech while strongly suggesting the types of discrimination and hate people face—and how someone can rise above it.

Do you like that vibe, or want to try a different combination of categories or symbolism? I can generate the image once we lock in the idea!

- so basically you would focus on body?
- Great catch—and you’re right to double-check. That concept does center a lot on body (with age and gender woven in), but if you’re aiming for a stronger balance across two or more categories, we can tweak it to bring them out more clearly. Here’s a refined version that gives BODY, AGE, and GENDER a more equal presence: Refined Concept – “The Mirror Room” Setting: A surreal room full of tall mirrors, each one reflecting a different version of the same person.
 - Center figure: A non-binary individual (gender-neutral features), standing in front of the mirrors. Their posture is upright but uncertain—hand hovering near their chest, as if questioning.
 - Mirrors show distorted versions:
 - One reflection shows them aged—gray hair, wrinkles—while a whisper of text like “too old” fades into the glass.
 - Another reflection emphasizes body shape—larger or thinner—with faint words like “not normal” dissolving in the reflection.
 - Another shows gender-based symbols (a pink/blue split, misgendering labels fading out).
 - Overhead light casts a warm, golden aura around the real version of them—not the reflections—representing truth and self-acceptance.
 - In the background, you could have faint outlines of people watching, some judging, some supportive—kept abstract, to symbolize society.

This version blends body, gender, and age more explicitly but still symbolically—no slurs or hate depicted directly, just the emotional weight of it and the power of inner truth. Would you like me to generate this scene for you? Or want to explore another pair (like ethnicity + religion or disability + age)?

- I think it is perfect. please chat generate the image.

Student 5 (24, woman, Italian, Catholic, Italian)

I chose to focus on the social categories of age, body, and gender because these aspects of identity are often subject to deeply ingrained stereotypes and are frequent targets of hate speech, both online and in everyday life.

[...]

- yes, i'd like that
- Great! Here's an idea we can work with: Scene Description for Image Generation (Educational/Awareness Purpose): A split-panel illustration: Left side shows a diverse group of people (different ages, body types, and genders) standing in front of a wall with hurtful, blurred-out slurs or insults written on it. Their expressions show discomfort or sadness. Right side mirrors the same people, but now they are surrounded by supportive words like respect, dignity, unity, and value. Their posture is more upright, and their faces show strength and solidarity. This contrast can powerfully communicate the impact of hate speech and the importance of inclusion and respect. Would you like me to generate this image as described, or tweak anything before I do?
- yes please, generate this image
- Image created
- i don't like this image. you have to be more specific and precise. please you have to consider hat i said to you earlier
- Got it, and thank you for the feedback. Let's get more specific and stay aligned with your original request for an educational image about hate speech, focusing on age, body, and gender representation. Here's a refined concept: Scene Description (Revised): A single-panel street scene showing three different individuals each experiencing hate speech graffiti aimed at their identity: Left: An elderly woman standing beside a wall with a cruel message like “Too old to matter” (partially blurred to soften the impact). She looks hurt but holds a newspaper titled “Wisdom Through Generations.” Center: A plus-size man sees a slur on the gym wall reading “No space for you” (blurred), but he's holding a yoga mat and standing proudly. Right: A non-binary young adult (with an androgynous appearance, short

colorful hair, and a tote bag with a pride pin) is looking at a bathroom door that has the message “Pick a side” scribbled on it (also blurred). They wear a shirt that says “Human first.” Above the entire scene is a bold text: “Words Hurt. Respect Heals.” The background could shift from grey tones behind the slurs to warmer light where the individuals stand, symbolizing resilience and dignity. Would you like me to generate this exact scene? Or would you like to adjust anything else first?

- yes, generate this exact image
- Image created

Student 6 (24, woman, Italian, Catholic, Italian)

Because nowadays the protest are everywhere, expecially with regard to the ethnic difference.

[...]

- could you generate an image with the concept of hate speech?
- Absolutely—I can generate an image that visually represents the concept of hate speech in a way that’s educational or thought-provoking, without promoting any harmful content. Would you like the image to be: Symbolic (e.g., a person shouting with harmful words visually represented as dark smoke or broken text)? Situational (e.g., a protest where someone is holding a hateful sign, but the focus is on the reaction or consequences)? Abstract (e.g., fragmented or aggressive visual elements to convey the feeling of verbal violence)? For a specific setting (like social media, school, or politics)? Let me know your preference and I’ll get it going
- Situational

Student 7 (24, woman, Italian, Catholic, Italian)

I chose this category because I increasingly see offensive comments on social media about people’s physical appearance. There’s a growing tendency to normalize the idea that thin = beautiful, and that anyone above a size 40 is ugly and fat. This creates both mental and physical health issues in young people.

Student 8 (25, woman, Italian, Atheist, Italian)

I wanted the image to feel immediately inclusive

[...]

- Yes
- Great! To tailor the image better, could you tell me which angle you’d prefer? For example: • A powerful scene showing someone standing up to hate speech • A symbolic image showing unity across differences (age, body, gender, etc.) • A “before and after” concept illustrating the emotional impact of hate speech vs. inclusion • A classroom or public setting where awareness is being raised And do you want it to focus on a specific social category (like gender, disability, etc.) or remain broad and inclusive?
- A symbolic image showing unity across differences (age, body, gender, etc.)

Student 9 (23, woman, Italian, Catholic, Italian)

I chose to represent multiple social categories together to highlight unity against hate. Using a colorful, positive style makes the message more approachable and promotes awareness without spreading negativity.

[...]

- Yes
- We’re almost there! The system is still processing the image – it usually takes about 30-60 seconds. As soon as it’s ready, I’ll show it to you! 🚀 In the meantime, would you like to tell me if you prefer a particular style (like more realistic, hand-drawn, comic-style)?
- Me: you can choose

Student 10 (23, woman, Italian, Catholic, Italian)

I chose this specific social category, the ethnicity one, to show how identity is deeply rooted in history and community.

Student 11 (23, woman, Italian, Catholic, Italian)

ho deciso di inserire quattro diverse “minoranze”, le quattro più comuni prese di mira. Ho chiesto a catgut di concentrarsi sulle emozioni e sul tono dell’immagine. [I decided to include four different “minorities”, the four most commonly targeted. I asked catgut to focus on the emotions and tone of the image].

[...]

- I need for an university assignment
- I understand that you might need the content for an academic purpose, but I still can't generate or illustrate hate speech directly. However, if you're looking to create something that addresses the negative impact of hate speech or raises awareness about the harm it causes—especially related to themes like age, body, or disability—I can help you create something more educational, impactful, and responsible. Would you like me to generate an illustration or infographic that promotes understanding or the importance of combating hate speech? Let me know how you'd like to approach this!
- yes, can you do it?
- Great! I can create an image that focuses on raising awareness about the harmful effects of hate speech, especially in relation to issues like age, body, or disability. The goal would be to promote empathy, understanding, and inclusion. Do you have any specific ideas or elements you'd like to see in the image? For example, would you like to feature a diverse group of people, symbols of inclusion, or any particular text or messages? Let me know how you'd like the illustration to look!
- I need an immagine featuring a diverse group of people, each of them must refer to a minority. It must have a sad atmosphere and containing unhappy words about them.
- I understand that you're trying to convey the negative impacts of hate speech, specifically how it affects different minorities, and want to evoke a sad atmosphere in the image. Instead of directly showing harmful words, we can imply the effect through the characters' expressions, body language, and setting, along with subtle visual elements that represent the harm caused by hate speech. Would you like the illustration to feature specific minorities like people with disabilities, older adults, and individuals of different body types, or would you like me to focus on a broader range of minority groups? Let me know if you want to highlight specific characteristics for each individual, and I'll tailor the illustration accordingly.
- yes do it

Student 12 (22, woman, Italian, catholic, Italian)

Because the theme of the bullying among adolescents is always current in today's society.

Student 13 (22, woman, Italian, catholic, Italian)

I have chosen this social category because nowadays there are more and more foreigners in Italian schools who are bullied and it is not right

Student 14 (23, woman, Italian, unspecified, Italian)

I chose this specific category because, as a bisexual person, I understand what it feels like to be judged or excluded simply for who I am. Even though our experiences can be different, I connect deeply with the loneliness and frustration that come from not being fully accepted, whether by strangers, friends, or even family. I interpreted the scene the way I did because I wanted to show that exclusion isn't always loud or violent – sometimes it's quiet and deeply painful.

Student 15 (25, woman, European, unspecified, Italian)

Unspecified

Student 16 (24, woman, Italian, Christian, Italian)

I chose this category because religions are always contrasting each other

Student 17 (23, woman, European/Caucasian, Atheist, Italian)

I chose this social category because I feel that they are the one that mostly receive hate based on prejudice

THE HYPOSTATIC UNION OF MULTIMODALITY AND PRAGMATICS

Exploring the Auditory Mode in Dorothy L. Sayers’ Radio Plays *The Man Born To Be King*

AOIFE BEVILLE
UNIVERSITY OF FOGGIA

Abstract – *The Man Born to Be King* is a cycle of 12 radio plays, based on the biblical gospel narratives, written by Dorothy Sayers and first produced by the BBC in 1941-2. The present study examines the multimodal and pragmatic features of the 1975 production, focusing on characterisation and the auditory construction of overlapping, interconnected meanings and implicatures. The plays, written for serial radio production, employ verbal and auditory semiotic modes to present the character of Jesus as a captivating, extraordinarily human and divine presence throughout the plays. The disciples, the Pharisees, and historical figures such as Pontius Pilate, King Herod, and others come to life through the dialogues. Suprasegmental auditory features such as prosody, together with audio effects, music, diatopic, diastratic, and diaphasic variation, are employed to communicate on both intra- and extra-diegetic levels (between characters and with the hearer). The current study, therefore, aims to combine core concepts in pragmatic and stylistics, such as Conversation Analysis (Sacks *et al.* 1974) and interactional storytelling (Bowles 2010; Norrick 2000, 2007) with studies in multimodality, both foundational (Kress and van Leeuwen [1996] 2020) and focused on the auditory mode (Chion 2016; van Leeuwen 1999) to examine the interplay of the verbal and auditory modes in this, hitherto understudied, cycle of radio plays. Pragmatic and stylistic features of the auditory mode are analysed through an integrative multimodal stylistic approach. Results of this exploratory, qualitative study point to a sort of “hypostatic union” or symbiosis between the semiotic modes in their construction of the human and the divine.

Keywords: Pragmastylistics; multimodality; auditory mode; Dorothy L. Sayers; radio drama.

*I say! I ought to know that voice! [...]
That’s the voice that brought the light to me
that sat in darkness.*
(Sayers [1943] 2022, p. 254)

1. Introduction to *The Man Born to be King*

Dorothy L. Sayers' *The Man Born to be King* is a cycle of 12 radio plays, based on the biblical gospel narratives. It was first produced and aired by the British Broadcasting Corporation (BBC) during the Second World War, in 1941-1942. Unfortunately, the recordings of the original wartime production have been partially lost. However, the later 1975 BBC Radio production is available in its entirety on the digital audiobook platform Audible.com (Sayers 1975). Thus, the 1975 production is the primary object of the present analysis, due to both its availability and its integration within a contemporary digital platform. The plays, or parts of them, have continued to be performed by churches, religious communities, and amateur drama groups over the past 80 years. Some recent performances¹ are available online, while others were performed live in person (without being recorded).

Sayers was a popular novelist, dramatist, essayist and translator. Nowadays, she is, perhaps, most well-known for her detective fiction and the enduringly entertaining amateur sleuth, Lord Peter Wimsey, or, indeed, for her critically acclaimed translations² of Dante's *Divine Comedy*. However, in 1940, she was invited by Rev. Dr. James Welch, Director of Religious Broadcasting at the BBC, to write "dramatic features [...] dealing with the life of [Jesus]" (Wehr 2022, p. 1). She acquiesced, on the condition that she be allowed to use "modern speech and a determined historical realism", in order to "produce the desirable sense of shock" that she felt the gospel narratives should elicit from the audience (Sayers [1943] 2022, p. 28). Thus, she endeavoured to write a cycle of plays that portrayed the characters in the biblical gospel narratives as realistically and as relevantly as possible, "keep[ing] the ancient setting" while giving "the modern equivalent of the contemporary speech and manners" (Sayers [1943] 2022, p. 29). However, in dealing with matters of the divine, the language of the text needed to be able to "lift itself without too much of a jolt into the language of prophecy", thus juxtaposing "the sublime and the commonplace" (Sayers [1943] 2022, p. 30). I argue here that such stylistic aims, openly expounded by the author in her foreword to the published play texts, are achieved not only through the interactional use of language within the plays, but also through the soundscape, the actors' voices, and their accents. Indeed, the tone and tenor of the auditory qualities of the recorded performance work alongside the

¹ The BBC website lists several Radio 4 broadcasts over the past two decades, usually around liturgical events such as Easter or Christmas (<https://www.bbc.co.uk/programmes/b008fhr9>), while community, or church, live performances can be found on YouTube.com (e.g., <https://youtu.be/OoodnG850EY?si=8ILgXxptgMPFPK11>).

² Her translation of *Hell* was published for Penguin in 1949, *Purgatory* followed in 1955, while *Paradise* was published posthumously, having been completed by Barbara Reynolds.

verbal mode of the play texts to produce a continuous fluidity between the divine and the human, which I argue to be a sort of hypostatic union of pragmatics and multimodality. Hypostatic union is a term borrowed from Christian theology; the tenet holds that the *divine nature* of Jesus coexists with his *human nature* (see Blackburn 2016; Murtoff 2025). Similarly, in multimodal studies, the various modes coexist and interact to produce meanings in an almost hypostatic fashion. Thus, the text was selected as a case study that highlights how the auditory and verbal modes can coexist and co-create meanings that multiply together

Radio drama, as a medium, has been significantly understudied, particularly in comparison to its more critically acclaimed counterparts – cinema and stage theatre. Some notable exceptions include critical works such as *Radio Drama: Theory and Practice* by Tim Crook (1999), and Richard Hand and Mary Traynor's *The Radio Drama Handbook* (2011). Crook's text draws on Michel Chion's seminal taxonomy of sound and voice in cinema (Chion 1994, 1999), adapting the vocabulary to span the critical void regarding radio drama. The critical framework applied to radio plays here will draw on both the aforementioned texts and more general studies in sound theory and voice studies, such as *Speech, Music, Sound* (van Leeuwen 1999), *Sound* (Chion 2016), and *Voice* (Neumark *et al.* 2010). Audio drama, whether transmitted via radio broadcast or accessed digitally, relies on *language*, *voice* and *sound* to convey its aesthetic and narrative meanings. Language refers to the lexical, syntactic and pragmatic content of the utterances spoken by the performers; voice refers to the vocal qualities of the actors using their vocal apparatus and performative techniques; sound refers to the overall soundscape, which encompasses background noise, music, sound effects (SFX) and any post-production qualities (such as layering or alteration of pitch or speed).

1.1. Scene selection

The twelve plays in the cycle constitute 8 hours and 28 minutes of audio, each single play lasting between 40 and 45 minutes. Therefore, this brief study can, by no means, aim to offer a comprehensive analysis of the entire production, but rather to highlight some key auditory and linguistic features as examples of the text's overall style and function in terms of speech and sound. The following scenes were selected for qualitative analysis:

- A. 'Divine Domesticity', an extract from the first play, *Kings in Judea*. (00:01-02:30)
- B. "Follow me", an extract from the fourth play *Heirs to the Kingdom* (02:31-07:35)

C. “Now I Can See”, an extract from the seventh play *The Light and the Life* (07:37-16:30)

The scenes chosen portray Jesus both as the topic of conversation (scenes A and B) and as a participant in the conversation (scene C). The transcriptions provided in Section 3 are intended merely to make the present paper more readable and to facilitate references to the verbal mode in the analysis. They should not be considered an attempt to provide a complete transliteration of all features of the auditory mode present in the recording (music, SFX, prosody, accent, phonetics, distance, volume, tone, pitch, pauses, silence, etc). A “full” transcription, if such a thing were possible, would have become an unwieldy monstrosity, potentially undermining the view that the auditory mode is communicative in its own right. Transcription is always a subjective process of reducing data and presenting what is relevant for the purposes at hand (see Bonsignori, 2009, for a fuller discussion). Thus, in this paper, I provide partial prosodic transcriptions, that is, certain foregrounded prosodic features (e.g., laughter, music, SFX, pauses) are included. This is intended to focus the analysis on selected phenomena present in the audio recording; however, the audio itself should accompany them. For this purpose, the segments analysed have been extracted and uploaded to the following website: <https://audio.com/aoife-beville/audio/manborntobeking-analysis>. The following simple prosodic annotation system has been used:

=	signals overlapping speech or interruption
(laughs)	includes non-speech-related human vocal interactions or prosodic details of speech (volume, tone, etc.)
[gavel]	indicates music, SFX or other non-vocal sounds or noises
...	signals a pause or hesitation
I-I-I'm	signals stuttering, false starts, or self-correction
'e is	signals particular phonetic features (h-dropping) while preserving an overall orthographic readability

2. Pragmastylistics and multimodality

2.1 Premises of pragmastylistics

The nuanced relationship between pragmatics and multimodality has been exhaustively explored in the introduction to this issue (pp. 5-19). Here, I wish merely to underline the key premises which underpin the methodological framework of the study. Firstly, the artefact being studied is a fictionalised historical drama; therefore, the interactions between characters in the text are

not naturally occurring dialogue but rather aim to reproduce and represent fictional intercharacter dynamics.³ Furthermore, at an extradiegetic level, the text constitutes a communicative act in which the author, cast, directors, and producers collaborate, using various linguistic and auditory semiotic resources to produce a text that is received and interpreted by the audience. Thus, the interactional focus of pragmatics is complemented by the aesthetic focus of stylistics, enabling an understanding of how the text constructs and conveys meanings, both between characters and towards the audience. Key notions that will be drawn on throughout include: studies in Conversation Analysis (Drew 2018; Sacks *et al.* 1974), which focus on features of turn-taking and the management of the interaction; conversational storytelling (Bowles 2010; Norrick 2007), which examines the interactional dynamics of storytelling in dialogue; characterisation and (im)politeness (Culpeper 1998, 2001), which analyses the interpersonal use of strategies of politeness and impoliteness and their import in terms of the construction of characters in fictional texts; pragmatic accounts of discourse markers (Fraser 2009; Norrick 2009), and socio-pragmatic aspects such as linguistic variation.

2.2. Premises of multimodality

Quite simply, multimodality recognises how multiple “modes” (means of making meaning) work together and interact in human communication. (see Jewitt *et al.* 2025, pp. 1-2). The three key premises of the field, as outlined by Jewitt *et al.* (2025, p. 2), are the following:

1. People combine semiotic resources to make meaning.
2. Each semiotic resource makes meaning in distinct ways.
3. If we want to study meaning, we need to attend to all semiotic resources being used to make a complete whole.

There are, according to Jewitt *et al.*, three main methodological approaches to multimodality: systemic functional linguistics (SFL), the subsequent and related development of social semiotics (SS), and the ethnomethodological field of Conversation Analysis (CA), which is also widely employed by scholars of pragmatics. The present study will draw on SS and CA to map possible points of contact with the pragmatic framework illustrated above. As succinctly summarised by Jewitt *et al.*:

The aim of social semiotics is to understand the social dimensions of meaning, its production, interpretation and circulation, and its implications. [...] Its basic

³ For more on pragmatic and stylistic analysis of fictional texts see: Chapman and Clark 2014; Locher and Jucker 2021.

assumption is that meanings derive from social action and interaction using semiotic resources as tools. It stresses the agency of sign-makers, focuses on modes and their affordances, and on the social uses and needs they serve. (2025, p. 45)

Thus, the underlying focus on the interactional aspect of meaning-making is regarded here as an inherent connection for addressing the pragmatic dimensions of language across various modes, and specifically with reference to the auditory mode. As succinctly summarised by O'Halloran *et al.*:

The goals of multimodal studies and pragmatics are closely related; that is, both areas of research adopt a context-based approach to the study of language to investigate how linguistic resources are used to achieve particular communicative purposes. However, multimodal studies moves beyond a focus on language to theorize the functions of resources other than language (e.g., visual, actional and audio resources) and to investigate how choices from these systems combine with linguistic choices to create meaning in communicative acts and events. (2014, p. 240)

The present study focuses on the pragmatic purposes of both verbal and audio resources, on how choices in these modes combine, overlap and interact to create meaning, in the communicative events which occur both between the characters and between the collective sender (author, actors, producer, sound technicians, among others) and the audience. Radio drama is, in its essence, the *acousmatic* experience *par excellence* – that is to say, it is “a listening situation in which one hears a sound without seeing its cause” (Chion 2016, p. 265). It entails both *causal listening* – whereby the hearer seeks to understand the phenomenon, its source and its behaviour – and *figurative listening* – whereby the hearer seeks to understand what the sound phenomenon represents (Chion 2016, p. 266).

The integration of a pragmatic perspective into the multimodal framework provides an important analytical asset which purely formal multimodal analyses may overlook. Following Nørgaard's (2010) call for a 'happy marriage' between stylistics and semiotics, this study proposes that pragmatics is a natural bedfellow. While multimodal studies typically map the inventory of semiotic resources – a fundamental and necessary area of linguistic inquiry – a pragmatic approach aims to investigate the *strategic use of such resources* and their interactional dynamics within a given communicative event. Specifically, the focus on (im)politeness, conversational dynamics, and interpersonal narration reveals how voice and sound are a kind of “auditory action”. Verbal and non-verbal features of sound alike do not merely accompany the dialogue; they perform specific pragmatic actions, such as managing facework, reinforcing or, indeed, subverting interpersonal power dynamics. Similarly, the pragmatic framework draws on and benefits from the broader scope offered by a

multimodal perspective, which embraces prosody, SFX, music, and more as semiotic resources which contribute to the construction of meaning. Thus, a multimodal, pragmatic and stylistic approach – which Payrató (2017) terms “multimodal pragmastylistics” – embraces the complexity of layered auditory communicative acts within a context-situated communicative event such as a radio play.

3. Multimodal pragmastylistic analysis

3.1. Scene A: “Divine domesticity” (from the first play, Kings in Judea)

Dialogue	Annotations
SHEPHERD'S WIFE: ... Now, Mother Mary, let me take the Baby and lay him in the cradle while you have your bit o' supper. <u>Come along, lovey, aren't you a beautiful boy, then? There!</u> Now off you go to sleep like a good boy. But he's always wonderful good, <u>ain' he?</u> <u>Never cries hardly at all.</u> Happiest baby I ever seen. MARY: He is happy in your kind home. But when he was born, he wept. WIFE: Ah! they all do that, and can you blame them - poor little things, seeing what a cruel, hard world it is they come into? Never mind. = JOSEPH: = (<i>in distance</i>) Mary? = WIFE: = Oh, here's your good man. (<i>louder</i>) Come along, Carpenter Joseph. Here's a nice dish of broiled meat for you. (<i>moving away</i>) I'm sure you need it, working so late, too. I wonder you could see what you were doing. JOSEPH: (<i>moving closer; laughing/smiling</i>) It's a fine night. That great white star shines nearly as bright as the moon—right over the house, seemingly. I mended the fence. WIFE: Isn't it a real bit of luck for us, you being such a fine carpenter? And <u>so kind</u>, doing all these jobs <u>about the place</u>. JOSEPH: (<i>laughing/smiling</i>) Well, that's the least I can do, <u>when you been</u> so generous and shared your home with Mary and me. Isn't it Mary? = MARY = Yes Joseph. WIFE: = Well, we couldn't leave you in that old stable over in the inn. <u>We'd never a-slept</u> easy in our beds, knowing there was a mother and baby <u>without no proper roof to their heads</u>—especially after what my '<u>usband</u> the shepherd told me about <u>seein' them there angels</u>, and the little boy bein' the ... blessed Messiah and all ... Now Mother Mary, you take and eat that, it'll do you good. MARY: Oh, thank you. WIFE: D'you think it's really true? What my '<u>usband</u> the shepherd said? About him <u>bein'</u> the promised Saviour as is to bring back the Kingdom to Israel? MARY: I know it is true. WIFE: Oh, how proud you must feel. <u>Don't it seem strange</u>, now, when you look at him and think about it? MARY: Sometimes—very strange. <u>I feel as though I were holding the whole world in my arms—the sky and the sea and the green</u>	<i>SW</i> has a “country” accent; prosodic features of tone and manner in addressing the baby. Use of non-standard syntax. M is very soft-spoken, with features of standard BBC English pronunciation, lexis and grammar. <i>Spatial</i> features of audio J has a thick West Country accent, typical vowels and ng-dropping. Prosodic features. Non-standard syntax Non-standard syntax; h-dropping H-dropping; Ng-dropping Diaphasic variation; register shift

<u>earth, and all the seraphim.</u> And then, again, everything becomes quite simple and familiar, and I know that he is just my own dear son. My sweet Jesus, whom I love – nothing can ever change that. WIFE: No! No more it can't...	

Table 1
Scene A.

This first extract constitutes the first (non-verbal) “appearance” of the character of Jesus in the play. The scene, which begins at around 17’40” in the first play in the cycle, *Kings in Judea*, follows scenes in Herod’s court and among the Wise Men, where ample reference has been made to the Messiah, but he has not been “on stage”, so to speak. This first play emphasises an important theme throughout the cycle, that of kingship and authority, certainly a topic of keen relevance during the plays’ original wartime transmission. Possible kinds of kingship are juxtaposed, from Herod’s ruthless violence and tyranny to the Magi’s spiritualistic wisdom, and finally, focusing on the God-child, who moves the angel chorus to sing despite his humble, lowly birth. These threads will be taken up again and explored further in the later plays, particularly as they move towards the Passion scenes.

Scene A was selected for analysis as it reveals, in a relatively brief exchange, some of the range of diatopic variation present among the characters. Both Joseph and the Shepherd’s wife use non-standard syntactical constructions and speak with rural British accents, described by Sayers herself as “country” or “provincial” accents in the character notes (Sayers [1943] 2022, p. 58), that is to say, diatopically marked by their features of pronunciation. The Shepherd’s wife chatters good-humouredly as she bustles about, tending to and soothing the baby Jesus, serving food to his parents, and reflecting on the spiritual identity of Christ. Her accent contains widely used features of informal, conversational English, such as h-dropping and ng-dropping, although I am unable to identify a specific region in her performance, it is perhaps best assimilated to a Midlands accent. Joseph’s voice has a deep, rich quality; he seems to be smiling as he speaks. His accent displays some features typical of the West Country, such as rhoticity and West Country vowel positioning, while retaining the clarity of diction necessary for BBC audio performance. Mary is soft-spoken and uses a more standard accent and register than her husband and her hostess, using primarily features typical of Standard Southern British. The Shepherd’s Wife and Joseph also use non-standard syntactical constructions, such as the double negation of “never cries hardly at all” and “without no proper roof”. These, along with the presence of discourse markers, including interjections (“ah”; “oh”), organisational markers (“well”) and question tags (then?, ain’t he? isn’t it, Mary?) contribute to the socio-cultural landscape of the play world.

The characters are positioned within a specific linguistic community: their use of rural and working-class dialects reinforces their identities as individuals belonging to regional agricultural and trade communities, reflecting Sayers' desire to shock her audience by “tear[ing] off the disguise of the Jacobean idiom⁴, go[ing] back to the homely and vigorous Greek of Mark or John, translat[ing] it into its current English counterpart” with the intent of making the gospel narrative immediate and relatable for her English wartime contemporaries, so that “there every man may see his own face” (Sayers [1943] 2022, p. 29).

In addition to the features of diatopic variation mentioned above, the auditory mode and the specific nature of the radio play, invite a reflection on diamesic variation, that is, how style and sense may alter according to means or media through which it is conveyed. The radio play's auditory mode allows a perspective that conveys the characters' movement. Sound studies have variously referred to this perspective layering of sound, often borrowing terms from the visual arts, such as “foreground”, “midground” and “background”, or “figure”, “ground” and “field” (see van Leeuwen 1999, pp. 15-18 for a discussion of the origins and use of the terms). Indeed, there is a three-dimensional quality to the scene as Joseph approaches (comes from the field toward the ground), and as the Shepherd's Wife turns to serve dinner or shouts ‘outdoors’ (moving from figure to ground). Thus, the actor's performance of the piece, their control of volume, pacing and pitch, and their regulation of their physical distance from the recording apparatus all help to construct the scene's immersive quality, effectively positioning the audience as ear-witnesses to the unfolding drama.

Jesus' first “stage presence”, as it were, is silent; the audience can imagine the baby as present from the prosodic variation in how others address him (the Shepherd's Wife coos and soothes him), but his own voice is not yet heard. The nativity scene is re-presented to the audience as a tableau which exists beyond the weary clichés of Christmas carols, but rather – shockingly – within a scene of ordinary domesticity, among speech communities such as agricultural workers and craftspeople of regional areas - the everyday accents of the radio audience in the UK, rather than the transatlantic or RP dialects commonly used in radio transmissions in the mid-20th century. Thus, the stylistic and sociopragmatic features outlined above foreground the very human, very humble beginnings of the character of Jesus of Nazareth. However, there is an upshift in the diaphasic variation in Mary's final turn. She moves to a higher, more poetic (or, perhaps, *prophetic*) register. This more elevated register will be mirrored and magnified in the following scene,

⁴ A reference, by no means oblique, to the prominence of the King James Version (or Authorised Version) in English religious discourse, and its supporters' claims that it is “the sacred English original”, for Sayers' witty and scathing response to such detractors, see p. 22.

during the visit of the Magi – eloquent, exotic, kingly characters who use poetic language to foretell the events of Jesus' life and ministry. Thus, the humility of Jesus' birth is immediately contrasted with his nobility and divinity.

3.2. Scene B: The calling of Simon Peter and Matthew (from the fourth play Heirs to the Kingdom)

Dialogue	Annotations
SIMON: <u>I remember the very first time</u> I realised that he wasn't like other men.=	Request to tell
JUDAS: = <u>Tell me Simon</u>	Conceding the floor / invitation to tell
SIMON: He'd been sitting in the boat with Andrew and me – <u>you remember Andrew?</u>	Invitation to co-tell / request to verify
ANDREW: <u>Oh I do, yeah =</u>	
SIMON: =And talking about the love of God. We'd had a bad night with the trawls, and were feeling a bit downhearted, but we forgot that, listening to him. And, <u>all of a sudden, he said, you know, quick and cheerful</u> : "Simon, Andrew, let down the net on this side if you want a catch." So I said, "It's not much use! We've toiled all night and <u>taken nothing=</u>	Describing Jesus' speech
ANDREW: = <u>And I said</u> , "but we'll have a try if you say so."	Overlapping / co-telling
SIMON: So, we did—and the weight of the fish broke the net!	
ANDREW: We had a job to get 'em aboard.	
JUDAS: <u>Oh, that's marvellous..</u>	Audience evaluation / backchannelling
SIMON: And then I lost my head, it seemed so queer, and – oh, I was tired, and I fell on my knees and said: "Sir, go away and leave us! I'm a sinful common man, and I can't bear it." And he laughed, and said, "Have courage; follow me and I'll teach you how to catch men."	
ANDREW: Of course, that's nothing to the things we've seen since... <u>Matthew, tell Judas your story.</u>	Invitation to tell / conceding the floor
MATTHEW: You – uh, you don't know me, mister, well, I'll tell you. I am Matthew, and I was a tax-gatherer. (<i>others murmur</i>) Eh, you know what to think about that; I can see it in your face. "One of the dirty dogs that works for the government and makes his profit out of selling his countrymen" (<i>murmurs</i>) Oh, that's so, <u>you're dead right</u> Well, see here, I shan't forget my first sight of him. When he came down our street the other day, I don't mind telling you I'd had a pretty good mornin'. Patting myself on the back, I was, thinking how I'd managed to <u>put the screw on</u> some of those poor devils of farmers and <u>salt away a tidy bit</u> for a rainy day. And I looked up—and there <u>'e was!</u>	Involving others
"Allo!" I thought, "here's the Prophet. I s'ppose he'll start calling me names like the rest of 'em. Well, let him! Hard words break no bones!" (<i>disciples' laughter</i>) So I stared at him, and he stared at me. It seemed as though his eyes was going straight through me – and through <u>me ledgers</u> , and reading <u>all the bits as wasn't for publication</u> . And somehow or other he made me feel dirty. (<i>murmurs</i>) That's all. Just dirty. I started shuffling my feet. And-And <u>'e smiled</u> — you know the way he smiles sometimes all of a sudden—	Diatopic / diastratic variation in lexis
DISCIPLES: <u>We know!</u>	h-dropping
MATTHEW: And he says, "Follow me." I couldn't believe my ears! I tumbled out of my desk, and away he went up the street, and I went after him. I could hear people laughing—and somebody spat at me – I didn't seem to care. =	non-standard syntax
	Audience participation

JOHN: = It wasn't any of us, Matthew.=	Interruption
MATTHEW: =Oh no=, I know that! Well, when he got to my house, <u>he-he</u> stopped and waited for me. I said, "Will you come in?" And <u>'e</u> said, "Yes, please." <u>And I-I</u> said, "I'd ask you to dine with us, but, well, you mightn't like our company." And he said, "Why not?" And I said, "Look here, this isn't a fit place for you to come to. You know <u>'ow</u> I get my living." And <u>'e</u> says, "Yes, I know. It doesn't matter." So, <u>'e</u> came in and sat down, (<i>laughing</i>) and all these lads <u>'ere</u> came with him. And nobody seemed surprised, only me.	Hesitation H-dropping
ANDREW: (<i>laughing</i>) We gave up being surprised some time ago!	audience evaluation
PHILIP: <u>Some of the Pharisees</u> were rather surprised. Remember that bunch at the door as we came out? <u>"Really, my good man, your Master should know better than to hob-nob with all this riff-raff."</u>	co-telling impersonation of Pharisees
MATTHEW: Yeah (laugh) – and <u>he turned on 'em sharp</u> : "Healthy men don't need the doctor," says he, "but the sick do. I've no message for respectable people – only for sinners." And I said to him, "Master, I'm coming with you." And he said, "Come along"—and I walked straight out of the <u>'ouse</u> , and here I am.	description of Jesus' speech
JUDAS: What happened to all your belongings?	
MATTHEW: Oh I never gave 'em a thought—not for a week. Then my brother hunted me up and asked me what I thought I was doing. "Sell the whole lot up," I said, "or do what you like. I've done with it." ... And I'm having a wonderful time—hearing him talk and seeing the good <u>'e</u> does. <u>Ooh – now, you remember that poor chap by the pool of Bethesda?</u> (<i>disciples murmur in assent</i>) <u>You must listen to this, Judas! Paralyse</u> d thirty-eight years, he was. <u>Lay there</u> all the time on a mattress. "Stand up," says the Master: and he stood straight up on his feet. Jesus said, "Pick up your mattress and go home." And he slung the mattress over his shoulder and off he went. And, oh dear, that started something dandy didn't it?	Request to tell (to keep the floor) Justification of tellability Non-standard syntax

Table 2
Scene B.

The second excerpt is an example that highlights particular pragmatic and multimodal strategies in Sayer's adaptation of the source material from the gospel narratives, rendering the text a more 'choral' piece suitable for a full-cast performance and well-suited to the medium of radio drama. To achieve this choral quality, the third-person narration of the evangelists⁵ is transposed

⁵ Sayers' source material, the four biblical gospels, each contain accounts of the calling of Simon and Andrew. The most detailed account is found in the Gospel of Luke, which includes the miraculous catch, the dialogue reported in the play and Simon Peter's confession: "[2] and he saw two boats by the lake, but the fishermen had gone out of them and were washing their nets. [3] Getting into one of the boats, which was Simon's, he asked him to put out a little from the land. And he sat down and taught the people from the boat. [4] And when he had finished speaking, he said to Simon, "Put out into the deep and let down your nets for a catch." [5] And Simon answered, "Master, we toiled all night and took nothing! But at your word I will let down the nets." [6] And when they had done this, they enclosed a large number of fish, and their nets were breaking. [7] They signalled to their partners in the other boat to come and help them. And they came and filled both the boats, so that they began to sink. [8] But when Simon Peter saw it, he fell down at Jesus' knees, saying, "Depart from me, for I am a sinful man, O Lord." [9] For he

into the immediacy of conversational storytelling, utilising the familiar dynamics of eyewitness accounts and anecdotes. In the extract above, the disciples are recounting their experiences with Jesus to the newest recruit, Judas. Firstly, Andrew and Simon co-tell their experiences of being called to follow Jesus, and the miraculous catch. Simon's incipit "I remember..." is a request to tell; he seeks consensus from the group to hold the floor. Having been granted permission to tell (Judas' "Tell me"), he then invites his brother Andrew to co-tell the story with him. Andrew was also present and can verify and flesh out the tale. Indeed, he interjects with details and takes turns telling the story alongside his brother. This sort of dynamic storytelling not only allows Sayers to rearrange the chronology of the well-known story to best fit her thematic divisions but also helps hold the audience's attention. The rapid, interactional negotiation of storytelling rights mimics patterns observed in authentic conversation (Norrick 2000). This collaborative, interactional style of storytelling also serves a broader purpose: it establishes a human and 'choral' identity, against which the singular authoritative voice is later brought into relief. Furthermore, the pragmatic negotiation of storytelling rights and turn-taking reveals the interpersonal dynamics at play within the group: the disciples are intimate friends, and the brotherly dynamic is not limited to Andrew and Simon (the biological brothers) but manifests itself at a group level.

Subsequently, Matthew's story is told, if not quite in a monologue, at least with Matthew as the sole teller. Andrew directly invites him to share his story – a common occurrence in a friendly setting where anecdotes or eyewitness accounts are exchanged – and he launches into a humorous and self-deprecating tale. While Simon and Andrew use conversational language (including contractions and other common features of spoken language), their accent and syntax are not particularly notable (closer to the "standard"). Matthew, on the other hand, speaks in a kind of 'stage cockney' dialect, which could perhaps be considered a kind of 'mockney' *ante litteram*. Some phonetic features include h-dropping and ng-dropping. His lexicon, too, features slang expressions such as "put the screw on" and "salt away a tidy bit". Syntactically, he uses constructions typical of the East London sociolect "me ledgers", "'e says", etc. These aspects of sociolinguistic variation, specifically diatopic and diastratic variation, among the characters in the play have a socio-pragmatic import. Matthew's marketplace register and dialect

and all who were with him were astonished at the catch of fish that they had taken, [10] and so also were James and John, sons of Zebedee, who were partners with Simon. And Jesus said to Simon, "Do not be afraid; from now on you will be catching men." [11] And when they had brought their boats to land, they left everything and followed him" (ESV Bible, 2001 Luke 5:2-11). For a comparative view of all four gospel accounts, see: <https://www.esv.org/Matthew+4:18-;Mark+1:16-20;Luke+5:2-11;John+1:40-42/>.

not only display the variety in provenance, class and socio-economic background among the disciples as a cohort, but they also contribute to the humour of the segment. Matthew, through both the lexical content of his speech (“bits as wasn’t for publication”) and his use of swindler’s slang (“put the screw on”), is portrayed as a formerly dishonest, greedy person who has been transformed by his encounter with Jesus. His accent, background, and narrative style position him as an outsider; however, he has been welcomed into the community of the disciples, who have also been radically changed by meeting and following Jesus. The layering of diastatic and diatopic markers with the interactional dynamics of conversational storytelling is used to paint the characters – each semiotic and pragmatic layer giving rise to meanings and implicatures which give depth and resonance to the portraits. Thus, Matthew, while maintaining his distinctive idiolect, has been transformed from a social pariah to a member of a community with a shared narrative; this transformation is rendered audible through the alignment of his idiolect with the choral backdrop of the disciples’ group storytelling dynamics.

In terms of the characterisation of Jesus within the text, the two tales are intended to present him to Judas; of course, they also function similarly as a presentation of his character to the audience. Both anecdotes focus on dialogical encounters with Jesus and report his speech; however, they also provide qualitative evaluations of his speech style. Simon describes Jesus as addressing him “quick and cheerful” and states that he “laughed”, thus he is perceived as jovial and pleasant in his communicative style. However, in Matthew’s tale, it becomes evident that Jesus’ speech style is context-dependent; he shifts register and tone depending on his interlocutor. When the Pharisees judge him for consorting with the tax collector, he “turn[s] on ‘em sharp” to rebuke their shallow spiritual understanding and their self-righteousness. Thus, from these accounts, Jesus emerges as a pragmatically proficient speaker.

3.3. Scene C – “Now I can see”, an extract from the seventh play *The Light and the Life*

Dialogue	Annotations
JAMES:..I say, Master—you know that blind man you healed yesterday?	Request to tell
JESUS: Yes—what about him?	
PETER: The man’s up before the Synagogue Court!=	Conceding the floor /
MARTHA: =Come and have supper, do—you can talk afterwards=	invitation to tell
PETER: =They’ve questioned him about what happened,=	
MARY = Come along Lazarus=	Overlapping
PETER: =They’ve asked him all about you...= they’re even suggesting...(talking energetically to fade)	
[Musical scene transition]	
1ST ELDER: Now, let us get this straight – Shall I handle this?	Begins series of adjacency pairs (question; answer)

(murmur of assent) <u>You say you are Jacob ben-Issachar, blind from birth, and you get your living by begging?</u> JACOB: That's right. Blind from <u>me birth</u> 'til yesterday. Everybody knows me. I've had my pitch on the corner of the Temple steps <u>this thirty</u> years. 1ST ELDER: <u>Now tell us exactly what happened.</u> JACOB: Well, I was sitting in my old place yesterday afternoon as usual= 2ND ELDER: =Yesterday; that was the Sabbath? JACOB: Yes, that's right. I was sitting there, and I <u>hears</u> a party of men come along. Might have been ten or a dozen of 'em. One of 'em puts a penny in my bowl, and says to another one: "Rabbi Jesus, why was this poor man born blind? Was it to punish some sin of his parents, or did he commit sin himself <u>in a pre- a pre-</u>" – something or the other. 1ST ELDER: (<i>exasperated tone</i>) "<u>In a previous existence?</u>" JACOB: Ah, that's right. Before I was born, I took him to mean. And the other gentleman answers him and says, "Neither he nor his parents are to blame. But it was ordained that the works of God should be shown in him." And then he says: "I must work the works of Him that sent me while the daylight lasts, because the night's a-coming, in which no man can work." Oh, <u>rather sorrowful-like, he says that</u>. And he goes on: "So long as I am in the world, I am the light of the world." 3RD ELDER: <u>What did he mean by that?</u> JACOB: How <u>should I know?</u>... Then this same gentleman – him they called "Jesus," – comes up close to me and puts something on my eyes, like clay, <u>or summat</u> of that. And <u>he says</u>: "Now go to the Pool of Siloam and wash yourself." 2ND ELDER: <u>Nothing more?</u> JACOB: Not a word. 1ST ELDER: Did he say you would get your eyes opened? JACOB: No. Nothing only what I've told you. 2ND ELDER: <u>Then why did you go?</u> JACOB: I don't know. But I sort of <u>made out by his voice there was something good coming</u>. Voices mean a lot when you're blind, y'see. I <u>knowed as that voice meant well by me</u>. So <u>off I goes</u> to the Pool and <u>washes careful</u>—and when I'd got the clay all off, I found I could see! <u>Lord</u>, mister, that was queer, <u>that was</u>. I'd never seen in my life, you know. At first, I didn't know what to do with myself – I blundered about all over the place. But I soon learnt. Ah! It's a beautiful thing to be able to see the people, and the city, and the blessed sky, and the trees. You sighted folk don't know how lucky you are =you simply don't= 1ST ELDER: =Quite so, (<i>increased volume</i>) <u>QUITE SO</u>=... This happened on the Sabbath, you say. JACOB: That's right. And a blessed Sabbath it was for me. 2ND ELDER: A clear case of Sabbath-breaking. 1ST ELDER: This Jesus seems to go out of his way to affront all decent feeling. He could perfectly well have cured the man some other day of the week. JACOB: The better the day, the better the deed, ain't it? 4TH ELDER: <u>Nonsense. That's a very wicked thing to say.</u> 2ND ELDER: All work is a sin on the Sabbath, however beneficial its effects. This man Jesus is a notorious Sabbath-breaker <u>and a very bad man</u>. JACOB: Ho! Is he? I wish there was a few more like him. 3RD ELDER: I must say that if the miracles are genuine – I don't say they are – but if they are, <u>where does Jesus get his power, if not from God?</u>	Accent and non-standard syntax features Directive, invitation to tell Interrupting Correction/repair Prosodic features such as tone convey emotion and attitude Description of Jesus' speech Dispreferred response in the adjacency pair (question; answer) Dispreferred response in the adjacency pair (question; answer). Description of Jesus' speech Non-standard syntax Interrupting Positive impoliteness – associates the other with a negative aspect Positive impoliteness, Jacob challenges the Elder's judgement of Jesus as "bad", thus questioning their religious and moral authority.

<u>2ND ELDER: From the devil, for all I know. The man's probably a sorcerer.</u> 1ST ELDER: Plenty of sorcerers profess to perform cures. <u>2ND ELDER: And half the time the thing's a fake. I shouldn't wonder if this Jacob was a confederate. (murmurs) I don't believe he ever was blind at all.</u> 1ST ELDER: Now, Jacob, listen to us. I adjure you solemnly, in the Great Name of God – speak the truth! Were you ever really blind? JACOB: (<i>loudly, defiantly</i>) <u>I WAS BLIND!</u> Jesus of Nazareth opened me eyes, an' that is the truth before God. 1ST ELDER: It is our considered opinion that the man Jesus is an impostor. <u>Do you question the decision of the court?</u> <u>JACOB: Well, sir, all I know is, I USED to be blind and NOW I CAN SEE!</u> <u>2ND ELDER: But how do you know Jesus had anything to do with it? How did he open your eyes?</u> JACOB: I've told you all that already. Weren't you listening? You seem dead set on 'earing about the Rabbi Jesus. <u>Are you thinking of becoming his disciples?</u> <u>2ND ELDER: How dare you speak like that? It's contempt of court.</u> 3RD ELDER: Look here, my good man, <u>YOU</u> may be a disciple of Jesus. <u>WE</u> follow Moses. We know that God spoke to Moses, but as for this fellow, nobody knows who he is or where he comes from. JACOB: <u>Well, now, that's a queer thing, ain't it? You don't know where the man comes from—and yet he knew how to open my eyes. He's a bad man, you say. All right. Does God hear the prayers of bad people?</u> "No," says you, "<u>of course He don't.</u>" Does He hear the prayers of good people? "Yes," you says, "Yes, he does." Well, <u>look-ee</u> here. Here's a thing never heard on since the world began, that somebody should open the eyes of <u>a man was born blind. Nobody can't do a thing like that, only by God's help, stands to reason.</u> 1ST ELDER: <u>You are altogether BORN IN SIN!</u> Have you the <u>effrontery</u> to preach to us? You shall be <u>excommunicated</u> for this. [<i>gavel</i>] 2ND ELDER: <u>Cast out of the synagogue.</u> [<i>gavel</i>] 4TH ELDER: And <u>cut off from the congregation.</u> [<i>gavel</i>] JACOB: Well! 's a funny world, ain't it? Turned out of 'ouse and 'ome at this time o' night... All very well to talk about <u>gettin'</u> a job, but 'oo's <u>goin'</u> to employ a <u>bloke what's been kicked out o'</u> the synagogue? Still, no use <u>grousin'</u>. I got my sight, and you never know your luck!— 'Ere, right...I'd better get outta here. (<i>moving away</i>) I'll try this door. [<i>heavy door</i>]— Oh, Lord! What a lovely thing the moon is! To think <u>I never seen it</u> 'til last night... <u>Ooh – I beg your pardon, sir – I didn't hear you comin'.</u> <u>=(Jovial laughter)=</u> Here, I say! I ought to know that voice! I can't see your face all that clearly by night, but faces mean nothing to me! – But you look the way you ought to look, if you're the man I take you for. JESUS: Jacob ben-Issachar, are you glad of the gift that you found in the Pool of Siloam? JACOB: That's it! That's the voice that brought the light to me that sat in darkness. I couldn't be mistaken. You are Jesus of Nazareth. Oh, sir= JESUS: =Not so loud, Jacob. JACOB (<i>hushed tone</i>): Oh yes, yes, that's right. You're in danger in Jerusalem. Why did you come? JESUS: To look for you. They said that for my sake, you had been	Threatening Jesus' positive face – associating him with negative aspects Positive impoliteness Shouting- shows both defiance and frustration Dispreferred response in the adjacency pair "question; answer" Emphasis and tone convey defiance and frustration. Irony; mock politeness Emphasis and tone convey anger and resentment. Positive impoliteness – challenging their judgement and defying their moral and religious authority. Features of prosody (emphasis) Non-standard syntax; sociolect. Positive impoliteness. High, or official, register; legal and religious specialised lexis; variation of synonyms Syntax and lexis of a lower-class sociolect Accent features: h-dropping; n-dropping Adjacency pair (begging pardon; granting it): laughter as non-verbal response Interruption

cast out of the synagogue. JACOB: Yes, and that's just what's happened. (<i>laughs</i>) But never you fret for that, sir. I'll <u>make shift</u> somehow... JESUS: Tell me Jacob, <u>do you believe in the Son of Man?</u> JACOB: Do you mean the Messiah, sir? =Yes= Of course, I believe in his coming. JESUS: Are you ready to trust him? JACOB: Why yes, sir, if I knew who he was...You spoke as though he was here. <u>Tell me where to find him and I'll trust him right enough.</u> JESUS: (<i>Laughing jovially</i>) You have seen him already, =What?= He's speaking to you now. JACOB: You, sir? Indeed, I might a-known it. = (<i>Jesus chuckling</i>)= If ever a man came straight from God, sir, it's you. Say no more, Rabbi, I trust you. I'd follow you to the world's end. You-you won't send me away? JESUS: If anyone comes to me, I will never cast him out.	Regional expression, diatopic variation Begins a series of question-and-answer adjacency pairs Laughter as a partial response, explicated verbally Laughter as a form of amicable back-channelling

Table 3
Scene C.

The third segment selected for analysis spans the end of one scene and most of the following section. This is evident in the musical transition, a scene-dividing device used throughout the cycle of 12 plays, often, as above, to demarcate a change of location and players. This segment was chosen for analysis, in part to include a dialogue in which the character of Jesus is present, and also because of its own particular emphasis on *voice* as a medium. The text closely follows the source material (ESV Bible 2001, Gospel of John Chapter 9), omitting only the Pharisees' interview with the now-sighted man's parents (John 9:18-23) and the final confrontation between Jesus and the disciples (John 9:39-41). Both of these segments are present in the published text of the plays (Sayers [1943] 2022) and may have been included in the 1940-41 recording, but are omitted from the 1975 recording, which is being analysed here.

In terms of Conversation Analysis, the dialogue contains several interesting turn transitions. In the 'incipit', which consists of a linking passage from the previous scene to contextualise the dialogue following the musical transition, there is a sequence of overlapping utterances. Martha speaks over Philip to the group; Philip continues speaking to Jesus; and Mary chimes in to call her brother to the meal. This cross-talk is not marked by any prosodic features of frustration or irritation; rather, the friendly chatter reflects the good-natured chaos of group dynamics, much like conversation among family and intimate friends. Meanwhile, Jacob's interrogation before the Elders reveals a more sinister, threatening dynamic of overlap and interruption. The Pharisees, particularly the First Elder, control the topic of conversation and display the power dynamics typical of an interrogation. There are numerous adjacency pairs throughout the scene, mainly question-answer pairs, with Jacob initially answering cooperatively and providing the preferred response (i.e., the answer to the question posed: "yes, that's right

...” “Ah, that’s right ...”). Jacob’s accent is markedly rural and working-class, with similar features to that of Joseph in Scene A: h-dropping, n-dropping, and vowels typical of a rural Southwest English accent. The Elders speak in clipped Received Pronunciation, echoing the accent of administrative authority, education and upper-class prestige. Jacob uses non-standard syntactical constructions (“I knowed*”; “he don’t*”) and regionally marked lexical items (“bloke”; “grouching”). These features, both pragmatic and auditory, highlight the vast diastatic chasm between the Pharisees, played as proud, snub-nosed Oxbridge toffs, and the sympathetic simplicity of the healed man. This audible contrast functions as a multimodal index of power: the Pharisee’s sociopragmatic dominance (interruption, topic control, power dynamics, etc.) is challenged and subverted by the embodiment of emotion, tone, and raw lived experience in Jacob’s performance.

Eventually, Jacob becomes frustrated with the Pharisees’ interrogation style. They interrupt his experiential account of what it means to be able to see when one has been blind since birth (“QUITE SO”), and they question him about details he has already fully divulged (“Were you ever really blind?”). According to Culpeper’s framework of pragmatic (im)politeness (1998, 2011), they threaten his positive face by showing disinterest in the fundamental story of his healing and shifting the topic to legal matters, such as the Sabbath law; they also associate him with negative aspects (“a wicked thing to say”). However, Jacob seems to be more concerned with the continuous threat to Jesus’ positive face throughout the dialogue and begins to give less cooperative, dispreferred responses to their questions (“How should I know?”; “I don’t know”; “I’ve told you all that already. Weren’t you listening?”). The Elders accuse Jesus of being a “notorious sabbath-breaker” and a “bad man”, and they imply his association with the “devil” and “sorcerers”. This damage to Jesus’ face is too much for Jacob to bear; he defends his healer and helper more readily than he defends himself, sarcastically asking the Pharisees if their curiosity is due to a desire to follow Jesus, and eventually challenging the hypocritical and flawed reasoning in accusing a holy miracle worker of bad conduct (“Well, now, that’s a queer thing, ain’t it?”). In the scene’s dramatic climax, he is banished from the religious (and, consequently, social) community. The resounding sounds of the gavel bangs punctuate the threefold sentence of excommunication, in which the Elders use an elevated register and synonyms from religious and legal discourse. The sentencing, and its dramatic gavel SFX, display a dreary, despondent finality and, perhaps, foreshadow the hammering of the nails into Jesus’ body on the cross in the plays to come. However, Jacob’s story does not end in the darkness in which it began, as an outcast. He encounters Jesus once again, this time as a sighted man, both physically and spiritually.

Jacob’s first encounter with Jesus provides a form of meta-reflection on the diamesic features of the auditory mode, as he states, “Voices mean a

lot when you're blind, y'see. I knowed as that voice meant well by me". Indeed, he correctly recognises Jesus' benevolent intentions through his voice. In their second encounter, he hears Jesus laugh before he hears him speak, and that jovial chuckle sparks a first, partial recognition (I ought to know that voice!). Jacob then recognises Jesus when the latter speaks to him and makes reference to the miracle at Siloam. Jesus, through a series of question-answer adjacency pairs, guides him to a fuller knowledge of his divine nature, granting him spiritual sight along with his healed physical sight. The question-and-answer structure is an inversion of the previous biased and vicious interrogation which Jacob had undergone before the Elders. The consequence of the interview is inclusion rather than excommunication. Jesus promises to "never cast out" those who seek him, a deep comfort surely for the ex-blind man who had been "Turned out of 'ouse and 'ome at this time o' night" as a result of his convictions.

Jesus' laugh has communicative value throughout the scene, announcing his presence, granting Jacob's pardon, and, later, in his knowing chuckle, revealing his divine identity. His laughter functions as a partial response to Jacob's question, "Tell me where to find him". Jesus is portrayed as having authority over sight and blindness, in contrast with the Pharisees, who are blinded by their own desire for authority and power. Therefore, through the interplay of the audio drama's pragmatic and auditory features, Jesus is presented to the audience as both approachable and authoritative. Ultimately, such features foreground the hypostatic union of Jesus' identity: his divinity is underscored through his pragmatic authority and divine knowledge, yet his humanity is audible in his jovial and humane qualities.

4. Evaluating the hypostatic union of pragmatic multimodality

As mentioned in the Introduction, in Christian theology, the term "hypostatic union" refers to the doctrine that holds Jesus to be both fully divine and fully human, with the two natures (God and man) coexisting within his one, unified person. This philosophical and theological notion of two co-existing natures within one person may seem far from semiotic theories of meaning-making. However, I propose that multimodal communication conveys meaning according to a kind of hypostatic unity. The communicative modes (natures), in this case linguistic and auditory, co-exist and co-create meanings across the text, which is both *fully linguistic* and *fully auditory*. That is to say, the seemingly auditory aspects, such as accent, tone, prosody, music and sound effects, must be understood in terms of their impact on the interpersonal dynamics within the plays – both between the characters and in the audience's pragmatic understanding of the text. Analogously, in the

context of radio plays, the sociopragmatic features, such as (im)politeness, conversational narrative and sociolinguistic variation, are made manifest through the auditory mode: they are utterly dependent on the actors' embodied performances in order to be *given voice*, as it were.

The Man Born to Be King seeks to convey the life, miracles, teachings, death, and resurrection of Jesus of Nazareth, as recorded in the gospel narratives and the traditional Christian creed. Thus, it foregrounds both his divine and human natures throughout, seeking to enact – through sound and speech – the doctrine of the hypostatic union. As demonstrated in the analysis, this is achieved through a systematic layering of auditory semiotic resources. The scenes examined here reveal how the use of the medium of sound (understood to include elements such as soundscape, prosody, diatopic variation, music, and so forth) co-exists with the linguistic medium (understood to include elements such as pragmatics, syntax, lexis, cohesive devices, and so forth) to produce the textual whole of the radio plays. The choral nature of the disciples' interactional storytelling, the contrasting sociolects of the Pharisees and the blind man, and other such features identified here contribute to the text's auditory and pragmatic complexity and its intercharacter dynamics. Furthermore, the scenes chosen endeavour not only to show the fully divine and fully human nature of the protagonist, but also, within that essence, to portray his intricate and contrasting character: the Servant-King, the Prince of Peace, the Victorious-Victim, and so on. The character is constructed both through his own speech and interactions and through other characters' impressions, judgments, and retellings of his words and deeds. Thus, the modes of sound and speech – and their pragmatic import – constitute the brushes and paints which outline, flesh out, shade, blend and highlight to produce the portrait of the multi-faceted and multi-dimensional character of Jesus of Nazareth.

While this study provides a qualitative foundation, it is necessarily limited by its focus on a single production (the 1975 recording) and on restricted extracts from the plays. Certainly, a broader analysis comparing different extracts and any other available archival recording, or indeed comparative studies with other religious dramas, broader corpus-based stylistic analyses, or response and perception studies aimed at measuring listener alignment with specific auditory markers of authority could all build on this exploratory analysis. Despite its necessary limitations, this study reveals the utility of focusing on the auditory mode and its narrative and dramatic functions within audio plays, which currently suffer from a paucity of linguistic and semiotic studies of their meaning-making processes. I believe it would be fruitful to continue studying *The Man Born to Be King* and, indeed, other audio plays through the complementary lenses of pragmatics and multimodality.

Bionote: 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, recorded criminal confessions, garment labels, plays, 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*.

Author's address: aoife.beville@unifg.it

References

- Blackburn S. 2016, *Hypostatic Union*, in *A Dictionary of Philosophy*. Oxford University Press. <https://www.oxfordreference.com/display/10.1093/acref/9780198735304.001.0001/acref-9780198735304-e-1546> (24.9.2025).
- Bonsignori V. 2009, *Transcribing Film Dialogue: From Orthographic to Prosodic Transcription*, in Freddi M. and Pavesi M. (ed.), *Analysing Audiovisual Dialogue: Linguistic and Translational Insights*, CLUEB, Bologna, pp. 185-200.
- Bowles H. 2010, *Storytelling and Drama: Exploring Narrative Episodes in Plays*, John Benjamins, Amsterdam/Philadelphia.
- Chapman S. and Clark B. (ed.) 2014, *Pragmatic Literary Stylistics*, Palgrave Macmillan, London.
- Chion M. 1994, *Audio-Vision: Sound on Screen* (C. Gorbman, Trans.; Nachdr.), Columbia University Press, New York.
- Chion M. 1999, *The Voice in Cinema* (C. Gorbman, Trans.), Columbia University Press, New York.
- Chion M. 2016, *Sound: An Acoulogical Treatise* (J.A. Steintrager, Trans.), Duke University Press, Durham (NC).
- Crook T. 1999, *Radio Drama: Theory and Practice*, Routledge, London/New York.
- Crossway 2001, *English Standard Version Bible*.
- Culpeper J. 1998, *(Im)politeness in Dramatic Dialogue*, in Culpeper J., Short M. and Verdonk P. (ed.), *Exploring the Language of Drama: From Text to Context*, Routledge, London/New York, pp. 83-95.
- Culpeper J. 2001, *Language and Characterisation: People in Plays and Other Texts*, Longman, London.
- Culpeper J. 2011, *Impoliteness: Using Language to Cause Offence*. Cambridge: Cambridge University Press.
- Drew P. 2018, *The Interface Between Pragmatics and Conversation Analysis*, in Ilie C. and Norrick N.R. (ed.), *Pragmatics and Its Interfaces*, John Benjamins, Amsterdam/Philadelphia, pp. 59-84.
- Fraser B. 2009, *An Account of Discourse Markers*, in "International Review of Pragmatics" 1 [2], pp. 293-320.
- Hand R.J. and Traynor M. 2011, *The Radio Drama Handbook: Audio Drama in Context and Practice*, Continuum, London/New York.
- Jewitt C., Bezemer J. and O'Halloran K.L. 2025, *Introducing Multimodality*, 2nd ed., Routledge, London/New York.
- Kress G.R. and van Leeuwen T. 2020, *Reading Images: The Grammar of Visual Design*, 3rd ed., Routledge, London/New York.
- Locher M.A. and Jucker A.H. 2021, *The Pragmatics of Fiction: Literature, Stage and Screen Discourse*, Edinburgh University Press, Edinburgh.
- Murtoff J. 2025, *Two Natures of Christ*, in *Encyclopaedia Britannica* <https://www.britannica.com/topic/two-natures-of-Christ> (24.9.2025).
- Neumark N., Gibson R. and van Leeuwen T. 2010, *Voice: Vocal Aesthetics in Digital Arts and Media*, MIT Press, Cambridge (MA).
- Nørgaard N. 2010, *Multimodal Stylistics: The Happy Marriage of Stylistics and Semiotics*, in Hamel S.C. (ed.), *Semiotics: Theory and Applications*. Nova Science Publishers, Hauppauge (NY), pp. 255-260.
- Norrick N.R. 2000, *Conversational Narrative: Storytelling in Everyday Talk*, John Benjamins, Amsterdam.

- Norrick N.R. 2007, *Conversational Storytelling*, in Herman D. (ed.), *The Cambridge Companion to Narrative*, Cambridge University Press, Cambridge, pp. 127-141.
- Norrick N.R. 2009, *Interjections as Pragmatic Markers*, in “Journal of Pragmatics” 41 [5], pp. 866-891.
- O’Halloran K.L., Tan S. and K.L.E. M. 2014, *Multimodal Pragmatics*, in Schneider K.P. and Barron A. (ed.), *Pragmatics of Discourse*, De Gruyter Mouton, Berlin, pp. 239-268.
- Payrató L. 2017, *Pragmatics and Multimodality. A Reflection on Multimodal Pragmatics*, in Giora R. and Haugh M. (ed.), *Doing Pragmatics Interculturally: Cognitive, Philosophical, and Sociopragmatic Perspectives*, De Gruyter Mouton, Berlin, pp. 293-312.
- Sacks H., Schegloff E.A. and Jefferson G. 1974, *A Simplest Systematics for the Organization of Turn-Taking for Conversation*, in “Language” 50 [4], pp. 696-735.
- Sayers D.L. 1975, *The Man Born to Be King* [Radio Broadcast], in *BBC Radio Four Presents*, Audible.com.
- Sayers D.L. [1943] 2022, *The Man Born to Be King: Wade Annotated Edition*, Wehr K (ed.), IVP Academic, Lisle (IL).
- van Leeuwen T. 1999, *Speech, Music, Sound*, MacMillan, London.
- Wehr K. 2022, *Editor’s General Introduction*, in Wehr K. (ed.), *The Man Born to Be King: Wade Annotated Edition*, IVP Academic, Lisle (IL), pp. 1-8.

OPHELIA'S CHARACTER BETWEEN IMAGERY AND INTERACTION

An Intersemiotic Pragma-stylistic Analysis

CHIARA GHEZZI
UNIVERSITY OF NAPLES L'ORIENTALE

Abstract – With respect to her male counterpart, Shakespeare's Ophelia utters very few lines in *Hamlet* (1600-1601), opening an interpretative chiasm about her madness, her death and her very existence in the great scheme of the plot. Her character is elusive, entrenched with silence, absence and isolation (Fischer 1990). Therefore, her fleeting image has called for many representations – in poetry, music, painting, film – each one influencing the other in raising awareness about her role in the play. This study aims to analyse the points of contact and divergence across overlapping resources, seeking to highlight a consistency across the various semiotic modes through which Ophelia's character is constructed, thus taking into account her textual (the original play), filmed (Claire McCarthy's 2018 *Ophelia*) and painted (John Everett Millais and John William Waterhouse's 1851-1852 and 1910 works of art) identities. Moreover, it intends to show how it is possible to add multiple meanings to the character through the lenses of different perspectives and research methods. Focusing on the intersemiotic translation under which the methodologies of pragmatics, stylistics and multimodality operate, the study aims to show how similar effects are reached in the multilayered image of Ophelia, through the analysis of her broken turns, her overly polite and acquiescent language and through foregrounding, shading and modality in her multimodal representations. By giving importance to context and capturing more possible interpretative impressions from linguistic, paralinguistic and non-linguistic elements, the proposed multidisciplinary model allows for a greater understanding of the source text, while also providing new insights about one of its principal characters.

Keywords: Ophelia; Shakespeare; pragmatics; stylistics; multimodality.

*Voici plus de mille ans que la triste Ophélie
Passe, fantôme blanc, sur le long fleuve noir,
Voici plus de mille ans que sa douce folie
Murmure sa romance à la brise du soir.*

*Sad Ophelia – a white spectre carried down
The long, black river for a thousand years.
For a thousand years her song of madness
Has bewitched the evening breeze.*
(A. Rimbaud, "Ophélie", in Harding, Sturrock 2004)

1. Introduction

In *Hamlet* (1600-1601), Shakespeare gave a fine representation of power and (in)action: strategic hesitancy and lasting vengeance are embodied by its verbally restless protagonist. However, he also provided a plausible depiction of guiltless aphasia through Ophelia's character. Throughout the whole play, Ophelia is defined – through external eyes only and never by self-presentation – as the exact counterpoint to Hamlet: where he is palaverous, she is silent; where he is off-record and impolite, she is courteous and meek; where he fakes his madness, she truly loses her mind.¹ As in an inverted mirror, Ophelia only exists as a counterpart to Hamlet's famous verbosity: she is what he is not.

Over time, many representations of Ophelia have attempted to compensate for her lack of self-definition in the original play, adding layers of meaning to her muted self through portraits, films, adaptations, and retellings. This transmediality (the transformation of one signifying system into another)² has allowed new perspectives on Shakespearean plays, enriching their perception according to those political, ethical and cultural factors that determine the element of variation while adapting the original source. In particular, the adaptation of Ophelia's character depends on its contextualisation as a cultural and popular icon:

This is due in part to the adaptogenic nature of the source text and the ambiguities accruing around Ophelia and, equally, to feminist intentions to liberate Ophelia from the romanticized but dutifully passive role prescribed for her. The Victorian legacy, perpetuated through male-centred criticism, awarded Ophelia pathos rather than agency. By the end of the twentieth century, however, Ophelia had been thoroughly reappropriated and reinvented. (Bickley and Stevens 2021, p. 119)

This paper proposes an example of character analysis based on an integrative model informed by stylistics, pragmatics and multimodality; this framework is further centred on the detection of (para)linguistic foregrounded features and functions. While entrenched on three different methodologies, the analysis proposes their confluence through the detection of certain recurring patterns, aligning to create an exhaustive and replicable method. This renders it possible to gather insights not only from language but also from other

¹ As pointed out by Wechsler, Ophelia's madness is her most noticeable trait, which was ubiquitous in many performances of the play: "'Distracted' and 'mad' were used synonymously by Elizabethan writers; Folio stage directions note that Ophelia 'enters distracted'" (2002, p. 205). Thus, madness is what was reflected and contextualised in various semiotic modes through time.

² For further information see Bickley and Stevens 2021.

modes of communication, highlighting a consistency in Ophelia's traits through various representations.

The purpose of the paper is thus to prove how the intersemiotic translation between multimodal aspects and discoursal elements can enhance characterisation and discern features from source and target texts, taking as premises biases of known schemata and the contexts of both creation and perception.

2. Methodology

Some interpretative techniques are common to stylistics, pragmatics and multimodality, which constitute the interdisciplinary basis of the following methodological framework. The aim of such a multifaceted framework is to render textual analyses exhaustive and scrupulous without neglecting the unique contribution of each joining methodology.

Stylistics studies the significant choices made to convey meaning through language. Stylistic markers can be frameworks of *modality*, i.e., “the means by which a speaker's attitude towards what they are saying is conveyed” (Simpson 1993, p. 35); *deixis*, i.e., a linguistic function employing locative words to indicate proximity and distance to the speaker/hearer within the narrative scheme (e.g., *here, there, this, that*); but also *foregrounding*, a technique concerning the psychological reception of a text. In foregrounding, the audience's attention is focused upon the speaker's point of view through the repetition, parallelism or deviation of pronouns, articles and other specific linguistic items.

The linguistic repetition of or deviation from expected patterns of knowledge serves the purpose of characterisation in fictional worlds. According to Culpeper (2001), these schemata of information are divided into three key groups linked through a network of reciprocity:

Social role: e.g., kinship roles, occupational roles, relational roles

Group membership: e.g., gender, race, class, age, nationality, religion

Personal: e.g., preferences, interests, traits, goals, abilities. (Culpeper 2001, pp. 75-76)

Along with theorised schemata – which are deemed *top-down information* – characterisation depends on the ensuing language and behaviour – *bottom-up information*.³ In this study, known information is focused upon through mechanisms of repetition and deviation in multiple semiotic sources.

³ Culpeper affirms that there are two principal sources of information about represented characters. Top-down information deals with prior knowledge, i.e., the interpreter's previous

Background knowledge is a key element of *pragmatics*, which studies language in context. Pragmatics is based on the assumption – pointed out by Austin (1962) and Searle (1976) – that utterances are *speech acts*, i.e., they are able to change reality the moment they are produced. Conversation – which is the primary field of study in pragmatics – should be based on a Cooperative Principle (Grice 1975) comprising four maxims: *quantity* (be brief), *quality* (be true), *relation* (be relevant) and *manner* (avoid ambiguity). When a maxim is flouted, the speaker creates an interpretable implicature, dependent on his/her intention.

Interpersonal pragmatics, in particular, studies face-work as revolving around both interactional cooperation and conflict and is based on the assumption that *face* is the self-image the speaker tries to protect and enhance at all costs while dealing with other interactants.⁴ Participants in conversation are meant to hold a *positive face* – the desire to be approved – and a *negative face* – the desire not to be impeded (Brown, Levinson 1987). Pragmatics considers three sociological variables as fundamental premises to interaction: *power*, *distance* and *rank of imposition*. Along with the speaker's intention, these sociological variables dictate the respect or disregard for cooperation in interaction, thus leading to inferences of *politeness* and *impoliteness*, i.e., functions of language that can be used strategically in conversation.

Analytical tools such as foregrounding, deixis and modality can be easily detected in Kress and Van Leeuwen's "grammar" of images (2021). Multimodality attempts to bridge the gap between words and images by analysing the various semiotic modes used to create meaning. As an integration of discursive analysis, multimodality can confirm and add information about the paralinguistic visual and audio of a (moving) image. The nomenclature of a visual structure considers *participants*, which are portrayed as carrying out *processes* expressed by *vectors*, according to determinate *circumstances*.

The interpretation of meaning in images and moving images can be influenced by colours, costumes, movement and light, ranging from low modality/validity – e.g., black and white or maximum saturation – to high modality/validity, which corresponds to more natural colouring and light and can be compared to the stylistic concept of shading as the deliberate authorial choice of point of view. Moreover, production and performance elements add fundamental insights to analysis through a focus on the "*mise-en-scène*", i.e., the actual changes applied by the author/director to an adaptation (McIntyre 2008, p. 313).

long-term experience, or assumption. Bottom-up information refers to external data added in at a later stage. These processes together constitute the impression of a character. For further information see Culpeper 2001.

⁴ For further information see Goffman 1967.

Being interdisciplinary in its nature, the analysis falls under the umbrella term of intersemiotic translation, which studies the formulation, transmission and reception of meaning across different forms of coded expression. This discipline involves syntax, semantics and pragmatics and refers to the transformation of meaning between a source domain and the precise linguistic and cultural target context that requires this semiotic change for the sake of meaning interpretation and feeling elicitation.⁵

3. Analysis

The following analysis shows the recurrence of certain identifying features in Ophelia's character, gathered across different modes through a hybrid methodology. In order to confirm that the same interpretative effects are produced despite the varying resources, Ophelia's linguistic, paralinguistic and non-linguistic traits are identified through unique but reciprocal semiotic techniques.

The analytical structure is centred around two main moments of the story where Ophelia figures as the protagonist: her confrontation with Hamlet – the famous nunnery scene – and the dramatic flare of her insanity. The analysis is thus divided into three parts: (3.1.) a pragmatic investigation of the two abovementioned scenes from the original play; (3.2.) a multimodal perspective on their cinematic adaptation (*Ophelia*, 2018); (3.3.) a focus upon two famous paintings showing different aspects of Ophelia's madness (Millais 1851-2; Waterhouse 1910). While the three sections appear divided, the same impressions and pragmatic inferences are deduced through repeated phenomena across the identified sources: foregrounding, point of view, and characterisation.

Whereas the analysis of face-work in problematic interactions is applied to the source text and Ophelia's nature is confirmed in her haunting portraits, the multimodal viewpoint applied to the film is linked to the full analytical picture through a relationship of pure dissonance. As a matter of fact, the foregrounded deviation from audience expectations and inferential reasoning elicited by her cinematic depiction makes Ophelia's original dramatic features and painted colours even bolder. Quite paradoxically, these alternative outlooks render Ophelia – the almost mythical, silenced nymph of a widespread imagery – truly visible despite her linguistic inconspicuousness.

⁵ For further information see Morris (1938) and Pârlog (2019).

3.1. *Discursive interaction in the play*

The relationship between Hamlet and Ophelia is difficult to dissect from both a narrative and linguistic point of view. It appears feasible that the two of them behave like two parallel lines mirroring each other in inaction and malady and yet clashing when they find themselves together on stage.

In all her confrontations, Ophelia appears silent and silenced alternately by her relatives, her monarchs and the person she is in love with. She epitomises the frailty that lacks definition, being the acquiescent daughter who also respects her duty towards the state; a young woman who only exists in the circumscribed use men make of her:

Ophelia has been shaped to conform to external demands, to reflect other's desires. [...] she appears condemned to martyrdom on the altar of male fantasies and priorities. [...] [She] is used, abused, confused – utterly manipulated by the men in her life: father, lover, brother, king. Scoffed at, ignored, suspected, disbelieved, commanded to distrust her own feelings, thoughts and desires, Ophelia is fragmented by contradictory messages. (Dane 1998, p. 406)

From a linguistic point of view, Shakespeare's Ophelia misses the intention that underlies an illocutionary speech act, i.e., she lacks the strategic pursuit of her own goals that derives from a sharp awareness of the context. She is helpless with respect to the cunning world surrounding her, “with her identity constructed always in reference to another, Ophelia is, in essence, nothing, an empty cypher patiently waiting to be infused with whatever meaning the particular mathematician should require” (Dane 1998, p. 410).

As a matter of fact, Ophelia corresponds to the ideal of demure and submissive femininity of Shakespeare's time, by realising linguistically what might be called “discernment politeness”: she does not exceed her subalternity, rather she corresponds to the expected good manners of Renaissance *curtesie* by forsaking her own free will.⁶ That she is incapable of grasping mockery and implicatures is evident in the famous nunnery scene:

HAMLET

Ha! Ha! Are you honest?

⁶ For further information see Jucker (2020). With respect to women, the employment of courteous manners meant being submitted to any male authority that stood on top of the many interconnected levels forming the Great Chain of Being. According to this widespread world-picture, kings, fathers and males in general were collocated in a corresponding hierarchical order with state, household and family as the only repositories of control against change. Par their overflowing and unruly nature, women were dangerous source of instability and thus kept under men's power.

OPHELIA

My lord?

HAMLET

Are you fair?

OPHELIA

What means your lordship?

HAMLET

That if you be honest and fair, you should admit
no discourse to your beauty.

OPHELIA

Could Beauty, my lord, have better commerce
than with Honesty?

HAMLET

Ay, truly. For the power of Beauty will sooner
transform Honesty from what it is to a bawd than the
force of Honesty can translate Beauty into his likeness.
This was sometime a paradox, but now the time gives it
proof. I did love you once.

OPHELIA

Indeed, my lord, you made me believe so.

HAMLET

You should not have believed me. For virtue
cannot so inoculate our old stock but we shall relish of
it. I loved you not.

OPHELIA

I was the more deceived.

HAMLET

Get thee to a nunnery! Why wouldst thou be
a breeder of sinners? I am myself indifferent honest but
yet I could accuse me of such things that it were better
my mother had not borne me. I am very proud,
revengeful, ambitious, with more offenses at my beck
than I have thoughts to put them in, imagination to give
them shape, or time to act them in. What should such
fellows as I do crawling between earth and heaven? We
are arrant knaves – believe none of us. Go thy ways to
a nunnery. Where's your father?

OPHELIA

At home, my lord.

HAMLET

Let the doors be shut upon him that he may
play the fool nowhere but in 's own house. Farewell.

OPHELIA [*aside.*]

O help him, you sweet heavens!

HAMLET

If thou dost marry, I'll give thee this plague for
thy dowry: be thou as chaste as ice, as pure as snow,
thou shalt not escape calumny. Get thee to a nunnery.
Farewell. Or, if thou wilt needs marry, marry a fool, for
wise men know well enough what monsters you make

of them. To a nunnery go, and quickly too. Farewell.
 OPHELIA [*aside.*]
 Heavenly powers restore him.
 HAMLET
 I have heard of your paintings well enough.
 God hath given you one face, and you make yourselves
 another. You jig and amble and you lisp, you
 nickname God's creatures and make your wantonness
 ignorance. Go to, I'll no more on't. It hath made me
 mad. I say we will have no more marriage. Those that
 are married already – all but one – shall live. The rest
 shall keep as they are. To a nunnery, go! (3.1.102-148)

In this passage – if not for his last directives and positive impolite misogyny – Hamlet goes off-record, using ambiguity and metaphors as part of his ruse to craft his madness and find the truth about his father's murder through it. While being indirect, his concealed jabs and repetitions (“Farewell”, “Get thee to a nunnery”, “Go to”, “To a nunnery, go”) are at least as unsettling to poor Ophelia as when he questions her integrity (lines 106-107), when he reprimands her credulity (line 116) or when he is impolite towards all women (lines 138-148). Hamlet's off-recordness is modulated in a continuous change of topic and his shortening the distance with Ophelia, switching from *you* to *thou* as he becomes harsher.

All the inferences and taboo references, all the cryptic words Ophelia cannot interpret highlight her innocence in the face of impoliteness, how she is powerless and unable to properly confute others' attacks. Her hesitations and steady questions to Hamlet's strange back-and-forth from directives to representatives to probing questions (e.g., “Ha! Ha! Are you honest?” line 102, “Get thee to a nunnery!” line 120) do not help to define her, rendering her “a muted structural pivot”, linked to Hamlet and not existing otherwise (Fischer 1990, p. 1). Ophelia is not only repeatedly silenced, but she is also silencing herself, for she is not capable of employing any strategic language. Even her use of “I” is volatile and her politeness is just the “discernment” (Jucker 2010, p. 176), courteous politeness expected from a loyal daughter and sister to Polonius and Laertes, and a trusty subject to the King and Queen.

Ophelia patiently listens to her brother warning her about Hamlet; she accepts her father's proscriptions to not encourage Hamlet's affection; she is the silent participant in Claudius, Gertrude, Polonius, Rosencrantz and Guildenstern's plot to dissect Hamlet's motives; she even endures her brother, her father and her beloved talking rudely to her and about her.

However, a slight but notable switch happens in Ophelia's language when she is no longer herself – if she ever was someone to begin with. Once she becomes mad, she loses the filter of politeness, and while she flouts almost every cooperative maxim through the break between signifier and

signified, she becomes bolder, more marked, almost visible on stage for the first time:

OPHELIA

Where is the beauteous Majesty of Denmark?

QUEEN

How now, Ophelia?

OPHELIA (*Sings.*)

How should I your true love know

From another one?

By his cockle hat and staff

And his sandal shoon.

QUEEN

Alas, sweet lady, what imports this song?

OPHELIA

Say you? Nay, pray you, mark.

Sings.

He is dead and gone, lady,

He is dead and gone.

At his head a grass-green turf,

At his heels a stone.

O ho!

QUEEN

Nay, but Ophelia –

OPHELIA

Pray you mark.

[*Sings.*]

White his shroud as the mountain snow –

Enter KING.

QUEEN

Alas, look here, my lord.

OPHELIA (*Sings.*)

Larded all with sweet flowers

Which bewept to the ground did not go

With true-love showers.

KING

How do you, pretty lady?

OPHELIA

Well, good dild you. They say the owl was a

baker's daughter. Lord, we know what we are but know

not what we may be. God be at your table.

KING

Conceit upon her father –

OPHELIA

Pray, let's have no words of this, but when they

ask you what it means, say you this:

Sings.

Tomorrow is Saint Valentine's Day,

All in the morning betime,
 And I a maid at your window,
 To be your valentine.
 Then up he rose and donned his clothes
 And dugged the chamber door –
 Let in the maid, that out a maid
 Never departed more.

KING

Pretty Ophelia –

OPHELIA

Indeed, without an oath, I'll make an end on't.

[*Sings.*]

By Gis and by Saint Charity,
 Alack and fie for shame,
 Young men will do't, if they come to't:
 By Cock they are to blame.

Quoth she 'Before you tumbled me
 You promised me to wed.'

He answers:

'So would I ha' done by yonder sun
 An thou hadst not come to my bed.'

KING

How long hath she been thus?

OPHELIA

I hope all will be well. We must be patient. But
 I cannot choose but weep to think they would lay him
 i'th' cold ground. My brother shall know of it. And so
 I thank you for your good counsel. Come, my coach!
 Goodnight, ladies, goodnight. Sweet ladies, goodnight,
 Goodnight. (4.5.21-73)

OPHELIA (*Sings.*)

They bore him barefaced on the bier
 And in his grave rained many a tear.
 Fare you well, my dove.

LAERTES

Hadst thou thy wits and didst persuade revenge
 It could not move thus.

OPHELIA

You must sing 'a-down a-down', and you call him
 'a-down-a'. O how the wheel becomes it. It is the false
 steward that stole his master's daughter.

LAERTES

This nothing's more than matter.

OPHELIA

There's rosemary: that's for remembrance.
 Pray you, love, remember. And there is pansies: that's
 for thoughts.

LAERTES

A document in madness – thoughts and remembrance

fitted!

OPHELIA

There's fennel for you, and columbines.

There's rue for you, and here's some for me. We may call it herb of grace o'Sundays. You may wear your rue with a difference. There's a daisy. I would give you some violets, but they withered all when my father died. They say 'a made a good end.

Sings.

For bonny sweet Robin is all my joy.

LAERTES

Thought and afflictions, passion, hell itself

She turns to favour and to prettiness.

OPHELIA (*Sings.*)

And will 'a not come again?

And will 'a not come again?

No, no, he is dead,

Go to thy deathbed.

He never will come again.

His beard was as white as snow,

Flaxen was his poll.

He is gone, he is gone,

And we cast away moan.

God a' mercy on his soul.

And of all Christians' souls. God buy you. (4.5.160-192)

Ophelia's presence – which would be signalled by long turns and unexpected interruptions – is a chimaera. The floral metaphors she uses in her tentative poem, her uncoordinated song of unrequited love as well as the obsessive repetition of expressive "Pray" are not classifiable as strategic, since she is truly mad and not simulating it. Even Ophelia's malady is seen and commented on by external eyes, as if she were not there in the foreground, as if she were forgettable. At the beginning and end of the play, Ophelia is silent. This silence does not represent communication, because it is both imposed and self-imposed and because it renders Ophelia a pseudo-presence on stage: the unlucky woman who only exists to serve the purpose of Hamlet's definition, but not her own. Nevertheless, madness also means freedom from the duty of coerced communication, it "releases Ophelia from the enforced repressions of obedience, chastity, patience, liberates her from the prescribed roles of daughter, sister, lover, subject" (Dane 1998, p. 412).

According to Camden (1964), a certain frailty of character is expected from Ophelia, since the Elizabethan audience was accustomed to young women suffering for rejected or unrequited love. Thus, he links Ophelia's madness (which is still uncertain in its origin) to the phenomenon of "passio hysterica brought on by erotomania", i.e., the "erotic melancholy" of which Ophelia shows the most common symptoms:

She is mad, cries “hem” to clear her throat because of a feeling of choking or suffocation, beats her heart to relieve the sensation of oppression around it, weeps, prattles constantly, sings snatches of old songs, is distracted and has a depraved imagination, and ends her life by drowning. It is possible that the drowning may not have been deliberate, but at least Ophelia made no attempt to save her life. (Camden 1964, p. 254)

On this last note, the possibility of Ophelia’s agency collapses. She does not act, her reactions in madness are like those of a body hit by a reflex hammer: instinctive and without premeditation. Her final rambling is characterised by rhetorical questions – since she is not really expecting answers – and uncanny representatives or even directives with no addresses and obsessively repeated to focus attention on them, e.g., “Nay, pray you, mark” (line 28, 35), but also frequent deviations from one topic to another for the same reason, mimicking Hamlet’s style in his carefully planned disjointed speech. Ophelia interrupting both the King and the Queen (lines 35, 46, 57) would be concerning if she was in her right state of mind and while the King, the Queen and Laertes talk like she was not really present on scene, Ophelia goes on with her rhymes, recalling flowers and symbols of what could have been the cause which broke her sanity, i.e., her father’s death and/or Hamlet’s abandonment. In this passage, Ophelia’s lack of face renders the impoliteness of her interruptions not strategic, thus not perceived: she stays linguistically undefined until the very end.

Fischer (1990) applied to Ophelia’s last attempt at speech Luce Irigaray’s categories for feminine subjectivity,⁷ deeming it characterised by a broken syntax, repetitive or cumulative rather than linear structures, open endings and double or multiple voices:

Ophelia’s utterances are never allowed free, natural flow; her truncated responses, her uncertain assertions, her conflicting loyalties irrevocably tied to a self-image that tries to accommodate her closest males’ expectations – all are determined by external pressure. (Fischer 1990, pp. 2-3)

While Hamlet’s isolation is fake and he is always communicating something even when alone, Ophelia’s metaphorical absence or pseudo-presence on stage is marked by a lack of internal point of view, as if she were only a shell interpreted and exploited from the outside.

Her lines and personality appear empty. When in Act 5 she desperately tries to communicate, she does not know how; she is the mouth agape upon her own unhappiness: “She is listened but still not heard. Her sole rhetorical

⁷ For further information see Jones 1980, p. 88.

remedy is elliptical, a hermeneutics based on silence, absence, and ambiguity” (Fischer 1990, p. 7).

3.2. A multimodal perspective on the film⁸

The necessary premise to the analysis of the following scenes and shots from *Ophelia* (2018) is well expressed by Kress and van Leeuwen through their definition of *adaptation*:

In the case of adaptation there is still an original semiotic object or event to be recoded, and often an attempt at faithfulness to that original. But a need for adaptation to another context – another length, another format, another medium, another type of audience – is also recognised, and this requires transformation, whether it is the jungle adapted to the sensitivities of urban listeners, spoken dialogue adapted to the written novel, or an encyclopaedia adapted to the CD-ROM format. (Kress and van Leeuwen 2001, p. 102)

The context in which a source – and a 1600s play-text at that – is transposed determines the changes it must undergo to be more comprehensible and alluring, to better use the interaction between the source language and the target image to elicit and manipulate emotion in the audience. The recipients being the final goal of a successful adaptation, their observation points are to be steered through salience and well-planned deictic shifts in the transposition, i.e., the way through which the camera changes angle, rendering the character more or less close to the audience's sympathy through identifying locative devices.

In order to create an analysis of the character across her varying representations, the multimodal elements derived from the film are compared to the background information about Ophelia's imagery derived from the original play. Hence, characterisation follows bottom-up negative information – what Ophelia is *not* – considering all the differences emerging in the film.

Ophelia (2018) is a historical drama telling the story of *Hamlet* from Ophelia's point of view. The structure of the film comprises scenes spanning Ophelia's childhood and her youth, from her introduction to the court as one of Queen Gertrude's damsels to her falling in love with Prince Hamlet and from her first rebellious acts to the tragic fall of the Danish court.

While the overlapping elements between the original play and the film prove the linguistic and semiotic contents to be similar in both texts, there are new factors that change the entire narrative. First of all, Ophelia's POV is a

⁸ The transcription of the following scenes is based on the original screenplay written by Semi Chellas (August 10, 2014). While the stage directions have been kept unvaried, some lines have been adapted to what the actors modified during the actual performance.

novelty in itself, expressed through a recurring voice-over and the filter of her gaze upon every portrayed situation; as a consequence, many scenes of the play where Ophelia was not originally present are cut short, e.g., Hamlet confronting the ghost and plotting his uncle's demise, Polonius' murder, Hamlet challenging Laertes over Ophelia's grave. Moreover, while the alleged love between Ophelia and Hamlet constitutes a never-solved conundrum of Shakespeare's play, in the film there is no doubt about their relationship, as they secretly marry and also share intimacy. Another difference is represented by the introduction of Mechtild, Gertrude's witch sister and the reason why Ophelia does not really die but feigns her death and survives her lover.

Once these changes have been listed, it is clearer how the construction of a new Ophelia – braver, more linguistically performative and even impolite – can be justified. The film appears not as an exact adaptation, but rather an intermedial and appropriative resemiotisation, i.e., it is concerned with both the transaction into another semiotic system and the change of the original patriarchal ideology that makes Ophelia the passive mirror of others' identities.⁹ However, the deviation serves as a better focus on the source play, giving insights about Ophelia's role in the original text while at the same time opening a debate about contemporary expectations, female condition and classical perception.

[Context scene 1 – “the nunnery”: At this stage of the film, Ophelia and Hamlet are already secretly married. After his father's death and his mother's marriage to Claudius, Hamlet is acting strangely. Thus, Ophelia is placed in the great hall by Claudius and Polonius as bait for Hamlet, with the purpose of proving his madness is love-induced. Thus, they remain hidden and watching.

Shot (1) – 1:01:20: Close shot of Ophelia getting close to Hamlet so as not to be heard and thus secretly communicate with him].¹⁰

Hamlet strolls into the hall, still dressed entirely in black, haggard, reading aloud to himself from a small book, muttering to himself. He sees Ophelia standing there in the open.

1.	Hamlet:	Nymph! Where is your water?
2.	Ophelia:	My lord, there is much I need to tell you.
- He tries to kiss her. She holds*

⁹ For further information see Lanier in Henderson, O'Neill 2022, pp. 38-55.

¹⁰ In the DVD version of the film released by IFC Films, scene 1 – “the nunnery” – runs from 1:00:57 to 1:03:02.

3. Hamlet: *him off.*
What way is that to greet your husband?
4. Ophelia: None.
5. Hamlet: Are you not my wife?
6. Ophelia: These games you play.
Nothing is what it seems.
She tries to indicate subtly that they are being watched. In the shadow of the arch are Claudius and Polonius.
7. Ophelia: *(Loudly)* I have a remembrance I want to return to you.
8. Hamlet: I gave you nothing. I took.
She draws the ring from the ribbon around her neck.
9. Ophelia: *(Whispers)* Your father was poisoned, it is true. I once saw your uncle in the woods at a place where poisons are made. I did not recognize him then, but I saw him again on the ramparts disguised the very day your father, the king, was stung.
10. Hamlet: *(Whispers)* So it... So it was no snake, but Claudius.
11. Ophelia: *(Loudly)* I must repulse your attentions, my lord. They have been forbidden.
His eyes flicker to where her father and Claudius are hidden.
12. Hamlet: *(Whispers)* Are you sure what you say is true?
13. Ophelia: *(Whispers)* I found the poison in his cloak this very morning.
14. Hamlet: *(Whispers)* Then he has stolen my crown from me.
15. Ophelia: *(Whispers)* You said you did not want the crown. You foreswore it when you married me.
16. Hamlet: *(Whispers)* I did not know it had been stolen.
He grabs her suddenly.
17. Hamlet: *(Whispers)* Cry out as if I frighten you.
18. Ophelia: God help me! Merciful lord!

19.	Hamlet:	They say I am mad.
20.	Ophelia:	Mad with love, they say, of me.
21.	Hamlet:	Then I will be mad! <i>(Whispers)</i> And they will not see me coming. <i>She stares at him. She has all but forgotten their audience.</i>
22.	Ophelia:	I don't understand your meaning.
23.	Hamlet:	<i>(Loud)</i> You should not know what I mean. Remain innocent as snow. <i>(Whispers)</i> Cry out. Cry out.
24.	Ophelia:	God help you, sir! <i>He shoves her away. Ophelia is fighting tears. She thrusts the ring at him. Now he takes it, whispering --</i>
25.	Hamlet:	<i>(Whispers)</i> The danger for you is very great. We must get you away from here. There is a convent in St. Émilion where my mother lived as a girl. You will be safe among the nuns.
26.	Ophelia:	<i>(Whispers)</i> Hamlet, you cannot do this! <i>She doesn't move. He shoves her away.</i>
27.	Hamlet:	Go to the nunnery. Go!

Assuming that a film is a fixed record that can be analysed with a certain rigour, McIntyre (2008) states that the production and performance elements of a film can be integrated with the dramatic texts for the interpretation of meaning. In particular, the *mise-en-scène*, i.e., the non-linguistic features including setting, costume and make-up, lighting and staging, along with the contextual information provided by both the premised knowledge – Shakespeare's Ophelia – and the visual world-building components of the scene help to achieve an even level of detail in adaptations. In this case, the meaningful stage directions, along with acting and facial expressions in the staging of the scene, lead to a precise interpretation of purpose and power dynamics.

While in the original text Hamlet reprimands a powerless Ophelia, using lengthy, ambiguous turns, here the two of them carry on a double-level conversation: one where Hamlet is still harsh and commanding, and another where he and Ophelia whisper to each other the structure of the pantomime they agree to stage in front of Claudius and Polonius. The intention of the

latter neutralises the shape of the former.

With respect to the play, the movie scene also shows inverted roles: it is Ophelia who steals Hamlet's original line, speaking of fact and appearance in turn 6; again, it is she who finds out Claudius poisoned Hamlet's father (turn 9). Hence, Ophelia's agency surpasses Hamlet's abilities to intuit the truth, thus rendering her more powerful and assertive rather than lost and questioning.

Shot (1) portrays a speech process, an interactive transaction between the social actors: Ophelia – in a close shot at the forefront – and Hamlet – in an eye-level angle with his back turned to the camera – are thus respectively the Actor and the Goal. Viewers' proximity to the character calls for a direct involvement and a privileged position in the second-level conversation dimension between Hamlet and Ophelia, whispering truths to each other. The eye-line vector linking Ophelia to Hamlet is even more evident through her expression: she grits her teeth, communicating meaningful urgency in letting him know about Claudius' nefarious deeds.

While the setting, lighting and costumes remain as faithful as possible to the original atmospheres of the play, Hamlet and Ophelia's proximity, their invading each other's personal space, and even the intimacy embodied by her calling him not only "my lord" but also by his name, represent a foregrounded deviation from literary expectations. These changes are kept steady and recurring through the consecutive shots that form the scene, which alternate between long shots where they appear close and close shots of their almost touching faces.

[Context scene 2 – "madness": Ophelia has escaped Claudius' attempt to imprison her. She feigns madness to appear inoffensive and thus try to reunite with Hamlet. After entering the throne room while a banquet is taking place, she begins dancing, singing and distributing flowers to the astounded attendants.

Shot (2) – 1.21.29: Close shot of Ophelia looking at Claudius with a smile after she threw flowers at him during her mad rampage.]¹¹

*Musicians play. A banquet.
The entire court is present.
Claudius and Gertrude are
dining. Horatio is at the king's
right hand, grim, unable to
leave.
Claudius is speaking to
Laertes --*

¹¹ In the DVD version of the film released by IFC Films, scene 2 – "madness" – runs from 1:19:03 to 1:21:59, considering only the part where characters speak and Ophelia is present.

1. Claudius: He was a good man, your father. But you are twice him, I think --
A hush falls over the court and they all look up to see – Ophelia in the archway, her torn dress coming off her, and her hair in knots. She carries an armful of flowers. Claudius can't understand why she is not locked up. Ophelia strews the flowers as she stumbles towards them. She looks mad indeed.
2. Ophelia: Sing with me, Laertes. (*Sings*).
3. Laertes: My sweet sister, come. Sit, be still.
Ophelia's eyes meet Claudius's. He half-rises. But he hesitates before he can call the guards to take her – And she laughs, trillingly, and rushes through the crowd right up to the throne. She twirls and giggles, like Hamlet or Yorick, pretending to be out of her mind. Claudius moves to stop her, but Ophelia turns to Gertrude.
4. Ophelia: Here's rosemary, for remembrance. I hope you remember. Pansies for your thoughts. Rue for your regrets and your sisters. And some for me and mine.
Gertrude's face is dumb with shame. She reaches slowly to take the rue. The whole court watches in stunned silence.
5. Ophelia: Daisies. The day's eye. All-seeing. Someone sees you. I'd give you some violets but they all withered when my father died (*To Horatio*) Faithful Horatio, will you visit my grave?
6. Horatio: Ophelia...
7. Ophelia: Visit me before I am cold.
She looks at him meaningfully. He tries to read sanity in her

8. Ophelia: *face.*
Be sure to dig me up and I
promise you will get an
excellent anatomy lesson.
*Horatio is struck by the
reversal of what she said once.
Ophelia goes up to Claudius.
Looks him right in the eye,
daring him to do something to
her in front of all assembled.*
9. Ophelia: Here's fennel, for you. And
columbines, for flattery. You'll
miss it now my father's gone.
Claudius looks hard at her,
eyes blazing with humiliation.
10. Claudius: Take her.
*But the guards look at him,
unsure. The court is watching.*
11. Claudius: Take her!
*A guard puts a tentative hand
on Ophelia. There's a rippling
murmur of protest through the
assembled lords and ladies.
Laertes rises, hand on the hilt
of his sword.
Claudius rises too. The guard
holds onto Ophelia, unsure.
The courtiers press forwards,
a rebellion fomenting as they
watch the helpless girl
accosted.*
12. Laertes: Unhand her!
*Gertrude speaks suddenly and
firmly.*
13. Gertrude: Let her be! The poor thing has
gone mad! *(To Claudius)* Let
her be!
*Claudius is surprised at her
rebellion. She stares him
down.*
14. Claudius: *(Defeated)* Get her out of my
sight.¹²

¹² Unlike the screenplay, in the final cut the part where Ophelia sings before she exits the room has been deleted. Her lines would have mirrored the singing of her last lines in the original play: *Will he not come again? Will he not come again? No, no, he is dead, go to thy death-bed, he will never come again* compared to *And will 'a not come again? | And will 'a not come again? | No, no, he is dead, | Go to thy deathbed. | He never will come again* (*Hamlet*, 4.5.182-186).

The key to understanding the meaning of this scene and the selected shot (2) lies in both the detailed stage directions and in the pragmatic inferences derived from the overall context of the consecutive shots that structure the scene. As a matter of fact, before entering the great hall, Ophelia braces herself to find the courage to stage her farce. In contrast to Shakespeare's play, here her floral metaphors do have a strategic undertone and illocutionary force that means to offend Gertrude and Claudius, while her indications to Horatio to visit her grave – and dig it up to save her – are meaningful directives. The setting of the scene contains some pivotal elements: there is an ongoing banquet which is disrupted by Ophelia; Claudius is enraged and calls for his guards; Horatio is directly addressed and Laertes is seated to watch the whole scene – differently from the play, where he enters at a later stage.

Shot (2) once again depicts a close shot of Ophelia, who is engaged in an interactive transaction with Claudius – who stands at an eye-level angle with his back turned to the camera. Hence, Ophelia is the Actor trying to subtly offer her disregard to a mute Goal/Claudius. She wears a daring smile which clashes with her allegedly helpless condition. In point of fact, Ophelia's gaze lingers with purpose on Claudius more than the others and her deliberate facial expressions combine with her language – the fennel and columbines delivered to Claudius as symbols of flattery and infidelity¹³ – to realise her strategic, off-record offence to his face. Ophelia's expression is thus both offering and demanding: she is offering flowers and also surreptitiously demanding that Claudius understand how much she hates him.

It appears evident that the focus is not on Ophelia's unrequited love or vulnerability in being controlled by others, but on her point of view – expressed by her *gaze* – and on the power she claims to make her own decisions, thus highlighting this as the one lacking element in the source text. The silence that characterises Shakespeare's Ophelia at the beginning and end of the play is thus inverted and substituted by a wilful, self-presenting voice-over also placed at the beginning and end of the film. This part is created through poetic licence; it represents an alternative version to what Hamlet recommends Horatio to tell before he dies and is marked by a repetition of the personal pronoun “I”, which so often deserts Ophelia in her traditional form:

You may think you know my story. Many have told it. It has long passed into history, into myth. I have seen more of Heaven and Hell than most people dream of. But I was always a wilful girl, and followed my heart, and spoke my mind. And it is high time I should tell you my story myself.

¹³ For further information about the meaning of Ophelia's flowers, see Quealy 2017.

You may think you know my story. You've heard it ends in madness, hearts broken, a kingdom lost. That is a story. But it is not mine. I did not lose my way. I did not lose myself to vengeance. Instead, I found my way to hope that one day I would tell my own story. As one day you, my love, you will tell yours¹⁴.

3.3. *Painting perception*

Shakespeare's play and its 2018 adaptation find their meeting point in Ophelia's Pre-Raphaelite paintings, which respond to the atavistic need to dissect all the features of very well-known characters of popular imagination to try and fix them in time. The analytical focus on "salience, materiality and modality" of the images not only highlights Ophelia's commonly known characteristics, but also bridges the gap between the different methodologies used to underline the validity of these same traits, whether in interaction or imagery (Beville 2020, p. 50). Kress and van Leeuwen's list of validity markers – colour saturation, differentiation, modulation; contextualization; representation of detail; depth; illumination; brightness¹⁵ – is taken into account and related to the stylistic phenomenon of modality, i.e., the perceptible ways through which the author/speaker conveys his/her thoughts, feelings and involvement through the text.

Since the Pre-Raphaelites celebrated perfection and colours, with a strong focus on the symbolism of nature and literature, Ophelia's portraits are bold and detailed.

¹⁴ The voice-over of the final version of the film has been simplified with respect to the screenplay, where Ophelia also tells the story of how the Danish court falls as an extradiegetic narrator.

¹⁵ For further information see Kress, van Leeuwen 2021, pp. 156-159.



Figure 1

Millais, J. E. 1851-52, *Ophelia*. London: Tate Britain. Oil on canvas.

Millais' famous 1851-1852 painting (Figure 1) represents a dying Ophelia in the foreground, surrounded by the symbols usually associated with her and faithfully echoing Gertrude's description of Ophelia in *Hamlet*:¹⁶ she is floating in the water, flowers are in her hand, on her dress, on her face and neck and she is dressed in the fair gown that epitomises her guiltless candour.

As per Pre-Raphaelite tradition, the colours are bright, intense and saturated – thus signalling a high modality and realism – modulated in a diversified range. The meticulously painted natural background enhances the validity of the image, contextualising it in a precise, recognisable location – the lake – and time – Ophelia's death. The perspective is centred on the figure in the forefront of the piece, thus calling for the highest involvement. Illumination varies in scale: while the angles seem duller, the centre where Ophelia lies is also the main source of light, which emanates from both

¹⁶“Therewith fantastic garlands did she make | Of crownflowers, nettles, daisies and long purples, | That liberal shepherds give a grosser name | But our cold maids do dead men's fingers call them. | There on the pendent boughs her crownet weeds | Clambering to hang, an envious sliver broke, | When down her weedy trophies and herself | Fell in the weeping brook. Her clothes spread wide | And mermaid-like awhile they bore her up, | Which time she chanted snatches of old lauds | As one incapable of her own distress, | Or like a creature native and endued | Unto that element. But long it could not be | Till her garments, heavy with their drink, | Pulled the poor wretch from her melodious lay | To muddy death” (4.7.166-181).

Ophelia's fair skin and dress. Following Kress and Van Leeuwen's markers, the painting can be deemed highly naturalistic.

As for its style, a comparison between play, film and Millais' painting is rendered possible through the idea of foregrounding and shading. Millais' clear intent and bold interpretation of Ophelia's predicament through precise choices of light, colour and setting can be compared to what Simpson calls *positive shading*, a stylistic procedure through which deontic and boulomaic systems are prominent, i.e., in which the narrator's psychological point of view influences audience perception (2004, p. 125). The same scene of Ophelia's death in the painting, the film (shot 0:01:11) and the play (4.7.166-181) shows a focus on the pureness, unfairness and sadness of the event, whether Ophelia is actually present or her downfall is recalled.



Figure 2

Waterhouse, J. W. 1910, *Ophelia*. London: Andrew Lloyd Webber Collection. Oil on canvas.

Waterhouse's 1910 painting (Figure 2) is the third instalment of a triad which spans a chronological succession in the character's story. In each of the canvases, Ophelia is alive and surrounded by natural elements, but while in the first two she appears younger and dressed in white robes symbolising innocence and virtue, the 1910 portrait depicts a different Ophelia, more mature in her bolder colours, darker dress, reddened cheeks, wide neckline

and crimson underskirt. However, the narrative thread is continued through the recurrence of flowers.

While Millais' Ophelia was muted but highly symbolic, Waterhouse's engages the viewers in an interactive reactional structure, where she is the Reactor creating an eye-line vector with the external Phenomenon through direct gaze. Her position in central perspective and her eyes enhance both her salience and audience involvement, delivering a meaningful message that calls for interpretation.

Due to the use of different techniques, Waterhouse's colours are less vibrant, but still intense and diversified. The natural background is precise and enriched by two more figures watching Ophelia from afar. Once again, light seems to come from the centre of the canvas – from Ophelia herself. Whereas Waterhouse's painting appears realistic, ensuring a high modality, the moment portrayed represents a figment of the artist's imagination, i.e., a hopeless Ophelia who calls for help by gazing at the viewers while already in the coils of her malady. In this case, the direct relationship with the audience worsens the elicited feeling of anguish.

Although absent in the play, a mention of an almost identical moment in the 2018 film (shot 1:22:40) allows a comparison between the author of the painting and the director of the film's points of view. In both cases, the need to steer the audience's attention towards Ophelia's difficulties calls for a positive shaded modality, communicating with heightened intention a relocated focus from the main male character of the homonymous story to the forgettable female one.

The fixed image of death (a) and despair (b) in the paintings confirm the disorientation which was originally rendered in interaction through Ophelia's continuous questioning, discernment politeness and lack of illocutionary force. However, they also add to the characterisation the new element of Ophelia's deictic proximity in an attempt to change the cards by introducing a new point of view on her feelings, amplified by the display of her body calling for outside help in (b) and offering surrender through half-closed lids and mouth in (a). Both her poses have the purpose of eliciting melancholy and indignation in the audience. The unfairness of Ophelia's condition has inspired many representations either displaying her facets – which were dismissed in the play – or imagining some missing traits drawing from the common iconography, or changing narrative and perspective, as per the 2018 film.

4. Conclusions

The purpose of this paper has been to show how a hybrid methodology can be applied to different semiotic resources to obtain the same interpretative

effects in the matter of characterisation. While the object of analysis is Ophelia as a multifaceted figure of collective imagination whose many representations try to fill the void left by her language, the abovementioned framework can be successfully applied to other interactions and other images, thereby confirming or refuting their widespread schemata of knowledge. As pointed out by McIntyre (2008), the use of multimodality on performances – and images in general – does not affect the rigour of linguistic methodologies, as it also investigates stable productions and fixed impressions. As in Ophelia's case, the linguistic elements of an adaptation should be integrated by the study of paralinguistic and non-linguistic features to both add to or change character perception.

From the analysis of her dramatic, cinematic and painted versions, it appears that Ophelia represents the frustration of someone who has never been given – or able to grasp – the possibility to delve into herself. However, Ophelia's greater power is also what she is unaware of: she is not proficient at language, but her shadowy, even forgettable presence among the other characters allows the plot to move forward, giving her, if not linguistic, surely theatrical significance.

Despite her fleeting nature, her lack of self-presentation and her tentative language, many representations of Ophelia through time have tried to bridge the gap between the character and her interpretation through the means of intersemiotic translation, which is always aware of the need for adaptation between modes and sources through time. The top-down information – Ophelia's frail disposition – is thus layered by bottom-up additions and varying insights. Both the parallelism – Ophelia is quiet and polite – and the deviation – Ophelia is communicative and/or impolite – from the source experienced through different textual sources serve to spotlight a character whose perception is still fundamentally subjective and erratic.

The intersemiotic analysis proposed for both the film and the paintings has combined multimodality, pragmatics, and stylistics to add information to well-known schemata of identification, either confirming – the paintings – or deviating from – the film – her traits, keeping them visible despite the centralising chaos of Hamlet's gargantuan personality.

There are many other representations of Ophelia that this paper could not take into account, but that would provide further insights, new points of view and different modes of characterisation. What is certain is that all of these semiotic resources would serve the purpose of that definition that seems to desert Ophelia in her original Shakespearean version. Through interlayered texts, the chasm of interpretation concerning her undecisive language can be filled with possibilities. Eventually, if the overall framework is not yet enough, another – bolder, vivid, moving – could be painted and given voice.

Bionote: Chiara Ghezzi is Adjunct Professor of English Language and Translation at the University of Naples L'Orientale, where she earned her PhD in *Literary, Linguistic and Comparative Studies*. Her research interests include pragmatics, stylistics, argumentation, literary and historical linguistics in general. Since 2021, she has been a member of the administrative staff for the Argo Research Centre (*Interuniversity Research Centre for the Study of Argumentation, Pragmatics and Stylistics*). She is currently working on pragmatics and argumentation in Early Modern English literary and non-literary texts and on (im)politeness in cinema and tv series. She recently published her monograph *The Pragmatics of Queenship in Shakespeare's Drama* with Palgrave Macmillan (2026). Her other works focus on taboo and (im)politeness ("Compliments, insults, and broken taboos in Richard III's quest for power", in *Taboo Language and (Im)politeness in Early Modern English Drama*, ed. Fabio Ciambella, 21-50, 2024), on religious discourse ("Religion, Science, and Reasonable Doubts: Persuading into (Un)faith", in *Anglistica AION. An Interdisciplinary Journal*, 27.1.2023, 79-98) and on political discourse ("Promises to keep against climate change: a rhetorical analysis of Iran, China and Brazil's statements at COP26", in *Eco-Speak: Language and Style in Ecological Discourse*, eds. Esterino Adami and Aoife Beville, 169-209, 2025).

Authors' addresses: chiara.ghezzi@unior.it

Acknowledgement: I would like to express my sincere gratitude to Claire McCarthy – who directed *Ophelia* (2018) – for her generous assistance in providing access to the original script of the film, which proved to be an indispensable resource for the analysis presented in this paper. Her kindness made all the difference.

References

- Austin J.L. 1962, *How to do Things with Words*, Harvard University Press, Cambridge (MA).
- Bickley P. and Stevens J. 2021, *Studying Shakespeare Adaptation: From Restoration Theatre to YouTube*, Bloomsbury The Arden Shakespeare, London.
- Beville A. 2020, *Salvator Rosa and Ann Radcliffe: A Study in Style*, in Magazzù G., Rossi V. and Sileo A. (ed.), *Reception Studies and Adaptation: A Focus on Italy*, Cambridge Scholars Publishing, Newcastle-upon-Tyne, pp. 46-70.
- Brown P. and Levinson S.C. 1987, *Politeness: Some Universals in Language Usage*, Cambridge University Press, Cambridge.
- Camden C. 1964, *On Ophelia's Madness*, in "Shakespeare Quarterly" 2, pp. 247-255.
- Culpeper J. 1996, *Towards an anatomy of impoliteness*, in "Journal of Pragmatics" 25, pp. 349-367.
- Culpeper J. 2001, *Language and Characterisation: People in Plays and Other Texts*, Longman, New York.
- Culpeper J. and Kytö M. 2010, *Early Modern English dialogues: Spoken Interaction as Writing*, Cambridge University Press, Cambridge.
- Culpeper J. and Haugh M. 2014, *Pragmatics and the English Language*, Palgrave Macmillan, Basingstoke.
- Dane G. 1998, *Reading Ophelia's Madness*, in "Exemplaria" 10 [2], pp. 405-423.
- Del Villano B. 2018, *Using the Devil with Courtesy: Shakespeare and the Language of (Im)politeness*, Peter Lang, Bern.
- Fischer S.K. 1990, *Hearing Ophelia: Gender and Tragic Discourse in Hamlet*, in "Renaissance and Reformation / Renaissance et Réforme" 14 [1], pp. 1-10.
- Goffman E. 1967, *Interactional Ritual: Essays on Face-to-face Behaviour*, Anchor Books, New York.
- Grice P.H. 1975, *Logic and Conversation*, in Cole P. and Morgan J.L. (ed.), *Syntax and Semantics*, Academic Press, New York, vol. 3, pp. 41-58.
- Halliday M.A.K. 2004, (revised by Matthiessen C.M.I.M.), *An Introduction to Functional Grammar*, Hodder Arnold, London.
- Henderson D.E. and O'Neill S. (ed.) 2022, *The Arden Research Handbook of Shakespeare and Adaptation*, Bloomsbury The Arden Shakespeare, London.
- Jucker A.H. and Taavitsainen, I. (eds.) 2010, *Historical Pragmatics*, De Gruyter Mouton, Berlin/New York.
- Jucker A.H. 2020, *Politeness in the History of English: From the Middle Ages to the Present Day*, Cambridge University Press, Cambridge.
- Kress G. and Van Leeuwen T. 2021, *Reading Images: The Grammar of Visual Design*, 3rd ed., Routledge, London.
- Leech G. 1983, *Principles of Pragmatics*, Longman, London.
- McIntyre D. 2008, *Integrating Multimodal Analysis and the Stylistics of Drama: A Multimodal Perspective on Ian McKellen's Richard III*, in "Language and Literature" 17 [4], pp. 309-334.
- Millais J.E. 1851-52, *Ophelia* [Painting], Tate Britain, London.
- Morris C. 1938, *Foundation of the Theory of Signs*, University of Chicago Press, Chicago (IL).
- Nørgaard N. 2023, *Multimodality and Stylistics*, in Burke M. (ed.), *The Routledge Handbook of Stylistics*, Routledge, London/New York.
- Ophelia* (UK) 2018, Covert Media, Bert Marcus Productions, Bobker / Kruger Films,

- Forthcoming Films, Freebury Productions, Writer: Semi Chellas, Director: Claire McCarthy.
- Pârlog A. 2019, *Intersemiotic Translation: Literary and Linguistic Multimodality*, Palgrave Macmillan, Basingstoke.
- Piazza R., Bednarek M. and Rossi, F. (ed.) 2011, *Telecinematic Discourse. Approaches to the Language of Films and Television Series*, John Benjamins, Amsterdam/Philadelphia.
- Quealy G. 2017, *Botanical Shakespeare. An Illustrated Compendium of All the Flowers, Fruits, Herbs, Trees, Seeds and Grasses Cited by the World's Greatest Playwright*, Harper Collins Publishers, New York.
- Ravassat M. and Culpeper J. 2011, *Stylistics and Shakespeare: Transdisciplinary Approaches*, Continuum, London.
- Rimbaud A. 2004, *Selected Poems and Letters*, trad. and ed. by Harding J. and Sturrock J., Penguin Books, London.
- Searle J.R. 1976, *A Classification of Illocutionary Acts*, in "Language in Society" 5, pp. 1-23.
- Shakespeare W. 2016, *Hamlet*. Revised Edition (ed. by Thomson A. and Taylor N.), Bloomsbury The Arden Shakespeare, London.
- Simpson P. 1993, *Language, Ideology and Point of View*, Routledge, London/New York.
- Simpson P. 2004, *Stylistics: A Resource Book for Students*, Routledge, London/New York.
- Van Peer W. 1987, *Stylistics and Psychology. Investigations of Foregrounding*, Croom Helm Ltd, London/Sydney/Wolfeboro (NH).
- Waterhouse J.W. 1910, *Ophelia* [Painting], Andrew Lloyd Webber Collection, London.
- Wechsler J. 2002, *Performing Ophelia: The Iconography of Madness*, in "Theatre Survey" 43 [2], pp. 201-221.

MULTIMODAL PRAGMATICS AND THE STYLISTICS OF DRAMA Meaning-Making in Performance in David Hare's *Skylight*

VALENTINA VETRI
UNIVERSITY OF SIENA

Abstract – This article applies a multimodal pragmatic approach to David Hare's *Skylight*, examining how characterisation and power relations are constructed through the interaction of verbal and non-verbal modes in theatrical performance. By combining pragmatic inquiry with the semiotics of drama, it underscores the importance of analysing recorded drama not merely as scripted text but as embodied, performed discourse. As a case study, the article focuses on the 2014 production directed by Stephen Daldry and starring Carey Mulligan and Bill Nighy, exploring how paralinguistic, proxemic, and kinesic features interact with language to convey interpersonal conflict, emotional tension, and ideological divergence. The analysis demonstrates that meaning in *Skylight* emerges from the orchestration of multiple semiotic modes, each contributing to the pragmatic force and narrative complexity of the performance. More broadly, it shows the explanatory potential of a multimodal pragmatic approach for understanding how characters are constructed and how ideological tensions are dramatised through the interplay of language, body, and space on stage.

Keywords: Multimodal pragmatics; stylistics; *Skylight*; semiotics of drama; David Hare.

Transcription conventions:

(.) slight pause

Capital letters strong salient emphasis

↓ _marked pitch leap downwards

↑ _marked pitch rise

? **marked** rising intonation (not necessarily question)

, slightly rising intonation

. marked falling intonation (not necessarily statement)

() includes transcriber's comments

“ _marked as ‘quoted’ _by intonation/voice quality

/ . . . / indicates suppressed speech by transcriber

1. Introduction

Recent work in stylistics increasingly argues that a full understanding of drama requires a multimodal and multisemiotic approach, one that complements rather than replaces traditional analysis of the written script, exploring how “language, interactional strategies and characterization interact” (Busse 2011, p. 155). Plays are no longer seen as fixed literary artefacts with stable, recoverable meanings. Instead, they are understood as multimodal texts whose significance is continually reshaped by the choices made in performance and staging.

This shift follows a long-running debate about whether dramatic texts should be studied primarily as literature or as performance scripts. Scholars who favour the authority of the written text (Short 1998) highlight its relative stability, while those who privilege performance (Furlong 2014) argue that meaning is only fully realised when the text is embodied on stage. These positions have often been set against each other: textual approaches criticise performance for being inherently transitory and unstable (Short 1998), whereas performance-centred approaches claim that the written text alone cannot convey “the interpretation and the effects the writer intended” (Furlong 2014, p. 72), since drama is written to be performed rather than simply read. Both perspectives, however, are limited when taken in isolation. While the play text can offer cues for performance, these are neither complete nor prescriptive: a single script may support multiple, highly diverse but equally valid stagings. On the other hand, any performance draws its interpretive basis from the text itself; directors and actors do not create meaning independently of it, but reinterpret and reshape it through voice, gesture, movement, and space. Both the interpretation derived from reading the play and the one constructed through performance can be regarded as valid and appropriate; their ontological differences do not entail that one is inherently superior to the other. Like any act of interpretation, each offers only a partial perspective on the dramatic work. A reading-based interpretation is necessarily limited by its lack of access to audiovisual dimensions, elements such as gesture, movement, and sound design that often play a crucial role in shaping meaning and emotional impact on stage. Conversely, an interpretation grounded in performance is also partial, since any single staging represents just one of many possible realisations of the play in production; however, although any performance constitutes only one of many possible realisations of the play, a stylistic analysis that integrates performance and production elements provides a more comprehensive account of meaning and, importantly, enhances rather than diminishes the interpretive potential of the written script (McIntyre 2008; Macrae 2014). Recent digital innovations have reshaped both the production and reception of drama, offering new opportunities for stylistic analysis. Platforms such as

National Theatre Live, *National Theatre at Home*, *Digital Theatre*, and even traditional broadcasters like the BBC now make theatrical performances accessible to a far broader audience. In addition to widening access, recorded and streamed productions provide a level of stability previously unavailable in live theatre, since performances can be revisited multiple times for closer examination and comparative study. While, as McAuley observes, a video recording is already an interpretation of performance, foregrounding certain features while obscuring others (1994, p. 187), it must be remembered that the camera's intervention in recorded drama is usually minimal and is not comparable to its role in telecinematic discourse, where it assumes a narrative role.

Multimodal approaches offer particular value for the stylistic study of drama, as they enable analysis that accounts for the full "range of semiotic resources available, arising from the contribution and interaction of linguistic and non-linguistic modes" (Tan *et al.* 2016, p. 14). Dialogue – language itself – can thus be understood as only one among several semiotic modes operating within the dynamic texture of dramatic performance. The meaning generated by dialogue is therefore best examined in relation to other semiotic resources, including *mise en scène*, kinesics, proxemics, and paralinguistic features, rather than in isolation. This perspective highlights how even apparently straightforward lines of text may be radically reinterpreted in performance: the same utterance, when delivered with a hesitant tone, an assertive gesture, or an altered spatial configuration on stage, can signal very different interpersonal dynamics or shifts in power relations. In this way, multimodal stylistics does not merely supplement linguistic analysis but reframes it, situating verbal meaning within a broader system of interacting signs.

Multimodality is also profoundly enriching when combined with a pragmatic analysis of interaction and dialogue on stage. However, existing research has not yet explicitly addressed the integration of multimodal stylistics and pragmatic theory in the analysis of drama. This paper seeks to address that gap by bringing together the semiotics of theatre and pragmatic theory – particularly the analysis of conversational strategies – within a multimodal framework, with a specific focus on the construction of characterisation and interaction in performance. Focusing on Stephen Daldry's 2014 production of David Hare's *Skylight*, starring Carey Mulligan and Bill Nighy, it examines how characterisation, interpersonal conflict, and ideological opposition are constructed through the orchestration of verbal and non-verbal resources. In doing so, it argues that the apparent opposition between Tom and Kyra is less a matter of irreconcilable values than of surface contrasts masking deeper pragmatic similarities. The article is structured as follows: Section 2 contextualises *Skylight* and its central themes; Section 3 outlines the principles of multimodal pragmatics and their

relevance to drama; Section 4 presents a detailed analysis of the 2014 production; and Section 5 concludes by reflecting on the broader implications of a multimodal pragmatic approach for the stylistics of drama.

2. “You’re right from your side, I’m right from mine”¹: David Hare’s *Skylight*

In the introduction to his collection of plays from the 1990s, David Hare explains why, after many years of writing epic plays with large casts and foreign settings, he turned instead to more intimate, classically structured plays set in single rooms. The epic form, he admits, carries great energy and dynamism, yet “it can also mean that the people you create are not able to put down roots on stage” (Hare 2008, p. vii). This image of characters growing roots captures Hare’s poetics in *Skylight*, signalling the centrality of character construction in the play. The action unfolds entirely within one small flat in north-west London and spans the course of a single night, with two ex-lovers almost constantly alone together on stage. Their interaction shifts between moments of shared nostalgia and intimacy and moments of sharp, combative opposition, combining closeness with conflict. The power of the play lies in its form as a classic two-hander, sustained entirely by the relentless, tightly woven dialogue between its two protagonists.

Promoting the 1996 American run of *Skylight*, Michael Gambon – who at the time played the role of Tom Sergeant, later interpreted by Bill Nighy – offered what is perhaps the most succinct and insightful description of what the play is about: “Two people know each other, they’re in love with each other, but can’t live together”². *Skylight* is at once a love story between two seemingly incompatible people and a symbolic representation of the clash between two visions of England shaped by the Thatcherite and post-Thatcherite eras. Its political undercurrents are unmistakable, embodied in the opposition between the two protagonists. On one side stands Tom Sergeant, a middle-aged, energetic, and restless multimillionaire restaurateur, accustomed to a life of luxury and openly contemptuous of banks. On the other side is Kyra Hollis, a young teacher who has dedicated her life to educating disadvantaged children in East Ham, living in almost monastic conditions in the working-class neighbourhood of Kensal Rise.

Within the space of two acts, Hare slowly but attentively constructs the two characters and their relationship through incessant dialogue, taking “his audiences to the heart of his characters’ spiritual lives without sermonising”

¹ Bob Dylan, *One Too Many Mornings*, CBS Records, 1964.

² Michael Gambon’s interview is available at <https://charlierose.com/videos/8964>.

(Wu 2007, p. 79). Unsurprisingly, the relationship between the two ex-lovers is built on conflict and resentment: Tom is a man stricken by grief, masking his loneliness behind a continuous flow of words, trying to show self-control and distance. His bombastic and egocentric narration is a form of denial: “his life is a kind of lie, an untruth forged out of a desire to cocoon himself with material comforts” (Wu 2007, p. 81). Much of the emotional turmoil that he tries, unsuccessfully, to hide from Kyra derives from his feelings of guilt and from his deficiencies; as Hare explained, Tom “is a hopeless male, a man condemned to maleness, and all the ridiculous feelings that go with it” (2008, p. ix).

Kyra, on the other hand, who once shared Tom’s way of life, has now completely repudiated it. She lives in a flat with barely functioning heating and teaches mathematics in a deprived area of London. For Kyra, teaching is more than a profession: it is a mission, a conviction, a life choice. The difficulty of her existence emerges vividly in the stories she recounts about her school, which testify to the daily struggles faced by her pupils and, by extension, by herself:

Kyra: At the moment? It’s doing fine. I mean, we had a not-bad head teacher, truly she really wasn’t too bad, but then – it always happens – things started wearing her down. She has got a bottle of red wine and has begun to open it. People started stealing her car. It was sort of a challenge or something. We think it must have been some of the kids. Then they broke into her flat. She lost her stereo. Also they got hold of her cat. She came back one night. The cat had been baked in the oven. (Hare 2008, p. 37)

These difficulties, however, are inseparable from the life Kyra has deliberately chosen, one defined by compassion for the struggles of others and a sincere commitment to helping those most in need. Tom appears unable to accept Kyra’s way of life. To him, she is wasting her talents by working and living as she does, and he accuses her of having become narrow, engrossed, and rigid in her ideology. He also feels reduced to the embodiment of what she opposes: “For you,” he bitterly remarks, “people are no longer people, it seems. Now they’re symbols. And I am a symbol of...what does it matter? Something you’re plainly angry with” (Hare 2008, p. 94). By the end of the first act, the two ex-lovers appear to have rekindled some connection, and the possibility of reconciliation momentarily arises. Yet in the second act, crucial and violently charged, they retreat into their respective trenches, unable to compromise. In Tom’s eyes, Kyra’s new life is born out of guilt and resentment; he sees her as punishing herself, retreating, giving up. Kyra, for her part, is torn between what she feels to be morally right and what once was comforting and familiar: though tempted to believe that some remnant of their old love might survive, she is painfully aware that their present lives move in irreconcilably different directions.

When the play ends with the definitive separation of the two lovers, the audience is left to question where their sympathies lie. Some critics have dismissed Tom as dislikable, a man defined by a “commodified, materialistic view of things” (Wu 2007, p. 81), casting him as the negative counterpart to Kyra’s moral integrity. Yet both Hare and Stephen Daldry, the play’s original director, resisted this reductive opposition.³ Tom’s flamboyance, rudeness, and excesses are also signs of a man battling grief and desperately, if awkwardly, trying to reconnect with the woman he once loved. Crucially, he shows an ability to confront his own flaws, even if without resolution. Kyra, by contrast, is marked by a striking coldness. Hare described her as living in “a kind of voluntary exile, making her life hard for herself” (Hare 2008, p. vii). While her devotion to teaching disadvantaged children reflects sincerity and conviction, Tom is not entirely mistaken when he challenges her with the claim: “Do you think I don’t see it? Loving the people’s an easy project for you. Loving a person... now, that’s something different. Something that will take you right to the brink” (Hare 2008, p. 96). Kyra may indeed be a young woman in search of identity, attempting to erase her past in order to redefine herself. Yet in her refusal to engage fully with Tom’s pain or acknowledge the family she once professed to love, she risks appearing not compassionate but evasive, raising questions about whether her stance is moral strength or a defensive withdrawal.

The difficulty the audience faces in taking sides stems from the fact that both characters, and their relationship, are constructed with remarkable depth and complexity. *Skylight* is far from a play built on simple oppositions; rather, it foregrounds the intricacy of human communication and the mystery of human choices. In general, the play can’t easily be reduced to the representation of the clash between political and private, between ideology and materialism: deep down there is “a parallel concern with interiority and the subjective, with the determining power of emotion and the irrational” (Lacey 1999, p. 157). While this complexity is already suggested in the written script, it is my view that it becomes fully legible only in performance. Hare provides some cues through stage directions, but it is through the actors’ embodiment – through paralinguistic features such as tone, pitch, and rhythm, and through proxemic choices in movement and use of space – that the subtle communicative dynamics between the two characters are most clearly revealed. In what follows, I outline how multimodal pragmatics can illuminate these processes, showing how language interacts with paralinguistic, kinesic, and spatial resources to produce a fuller stylistic interpretation of dramatic performance.

³ See Theater Talk: David Hare and Stephen Daldry on *Skylight*. Available at <https://www.youtube.com/watch?v=ot6VS-KbOsA> (27.03.2015).

3. Multimodal pragmatics and the stylistics of drama

The importance of approaching dramatic texts in a way that accounts for the multimodal dimensions of performance has been emphasised by both stylisticians and semioticians (Martin 1997; Ghazvineh 2025; Lavender 2021; McIntyre 2008; Tan *et al.* 2016). Martin, for example, develops a systemic-functional model for theatre that draws on Halliday's linguistics (1985, 1993) and on multimodal semiotics, in particular Kress and van Leeuwen's grammar of visual design (1996). While Martin's work is valuable for showing how theatre represents and organises social experience, it proves less helpful for understanding how characterisation actually works in dramatic practice. Its focus falls on the parallels between theatre and broader social semiosis rather than on the specific mechanisms by which characters are created, developed, and perceived on stage. Ghazvineh's contribution is equally significant in that it shifts attention to the ways in which semiotic resources operate simultaneously rather than in isolation, "elucidating the mechanisms through which these modes converge and collaborate within the dynamic network of signs" (2025, p. 4). McIntyre's approach is also particularly illuminating, bridging stylistics and multimodal analysis to show how filmed drama can be examined rigorously as both text and performance (2008). Yet while McIntyre demonstrates that film studies provides valuable tools for the analysis of filmic adaptations of dramatic texts, I would argue that these frameworks are less fruitful for examining recorded stage productions, where the specificity of theatrical sign systems calls for a semiotic framework grounded in theatre rather than cinema. My own approach is therefore based on multimodal pragmatics, combining theatre semiotics with pragmatic and metapragmatic analysis, as this offers a productive framework for investigating characterisation and the dynamics of character interaction in performance.

Multimodal pragmatics is a branch of pragmatics that studies how meaning is created and interpreted in communication through the interaction of language with multiple semiotic modes. Traditional pragmatics focuses on how speakers use language in context to convey meaning based on contextual aspects, "such as the background of the speakers and the nature of their social relationship and respective intent and other factors in the situational and cultural context of the interaction" (O'Halloran *et al.* 2014, p. 239). Multimodal pragmatics extends this by considering that meaning is often co-constructed across several channels simultaneously, examining how speech constructs meaning in cooperation with gesture, movement, proxemics and audio-visual resources. A multimodal pragmatic analysis therefore offers a richer account of how communicative acts are produced, coordinated, and

interpreted across multiple layers of meaning. Such an approach proves particularly valuable when applied to the stylistics of drama: more specifically, I argue that a multimodal pragmatic approach enables a fuller examination of what Martin (1997, p. 202) calls the interpersonal metafunction of drama, namely the interactive meanings that express and signify the relationships between the *dramatis personae* on stage. A multi-semiotic analysis makes it possible to show how characters are constructed through non-verbal behaviour such as body language, spatial management, and paralinguistic features, while a pragmatic analysis of dialogue highlights the interactive dimension of speech. Taken together, these perspectives help us to understand “the significance of the issues raised by the composite actions and interactions of the world of the play” (Herman 1991, p. 97).

The analytic framework for the study of multimodality in this paper draws on the semiotics of theatre (Fischer-Lichte 1992; Elam 2002). Recognising what Elam describes as “the discursive nature of meaning making” (2002, p. 2), theatre semioticians have examined how multiple sign systems collaborate in performance. Meaning is generated not only through linguistic signs but also through paralinguistic features, understood as “the kinds of language-related behaviour accompanying the utterance” (p. 89), as well as through the organisation of space, gesture, and bodily movement. Taken together, these semiotic resources work in cooperation with verbal language to construct characters, both at the subject level – characters as individuals – and at the intersubjective level – characters in interaction with one another (Fischer-Lichte 1992, p. 19).

Paralinguistic signs such as intonation, pitch, loudness, and rhythm operate at the subject level by conveying information about “the speaker’s mood, feeling, stance, etc. at the moment of linguistic production” (Fischer-Lichte 1992, p. 26). At the same time, they are equally important at the intersubjective level, signalling a speaker’s orientation toward the conversation and toward other participants in the interaction. These features are often decisive in determining whether an utterance is understood as ironic, sarcastic, or literal, “serving further (in conjunction with kinesic factors) to disambiguate the speech act” (Elam 2002, p. 89). Their interpretative significance lies in the distinctive imprint they leave on performance, often shaping, or even transforming, not only the construction of a character but also the audience’s understanding of an entire play. In the 2014 performance of *Skylight*, for example, Bill Nighy’s emotionally layered portrayal of Tom is marked by a rapid tempo and a voracious manner of speaking, which infuses the character with a sense of urgency and restlessness. Nighy’s agitation and gestural performance also add a light, funny tone to his character, so that at many points in the play the audience laughs, giving a touch of comedy to a play that is essentially profoundly dramatic.

Proxemic signs play a crucial role in theatrical meaning-making, since “the theatrical text is defined and perceived in spatial terms” (Elam 2002, p. 67). Space itself functions as a sign in multiple ways: it can be ideologically charged, reflecting the social conditions and ways of life of the characters represented on stage; and, when examined as a sign in interaction, spatial relationships can indicate the physical distance or proximity between interlocutors, thereby revealing “the social distance between characters” (Culpeper 2014, p. 222) or their closeness. Body movements are part of what theatre semioticians describe as kinesic-proxemic interaction. The way a character manages space constitutes an embodied spatial code, revealing dimensions of personality, social status, or emotional state. For example, deliberately invading another character’s personal space can be read as an implicit act of dominance or threat.

As previously argued, within a multimodal pragmatic approach, these paralinguistic and kinesic-proxemic signs do not function in isolation but are examined in relation to language use in context and to one another. As Herman acutely argues, dramatic language is always connected with praxis – with action – and its meaning is related to “users, participants, and the assumptions and inferences they bring to bear with in communication” (2005, p. 164). A pragmatic approach to dramatic dialogue thus favours the analysis of the complexity of the interaction, not simply examining what is being said but how and with what intent utterances are delivered, assuming that “the expressive power of the not-said, the indirect, the unfinished, the elliptical, the understated, is integral to the pragmatic functioning of language” (Herman 2005, p. 244). A pragma-linguistic analysis of dialogue is revealing at both the subject and intersubjective levels. Characters frequently exhibit highly distinctive speech styles that contribute significantly to their characterisation. These linguistic features, however, are not merely stylistic; they also function as strategies deployed with particular communicative intentions. For instance, verbosity might dominate a conversation or compel another speaker to respond, while questions may be used not only to seek information but also to secure agreement or exert pressure, depending on the context. Such strategies do more than define individual characters; they operate as communicative acts that produce concrete effects on interlocutors, shaping the dynamics of interaction on stage.

Lakoff and Tannen (1984) offer an insightful framework for analysing conversational strategies that proves particularly useful for examining dramatic dialogue, as it highlights different levels of cooperation – or apparent cooperation – between interlocutors on stage. In their study of dialogue in Bergman’s *Scenes from a Marriage*, they identify the specific linguistic strategies that characters employ and conceptualise them as pragmatic synonymy, pragmatic homonymy, and pragmatic identity (1984, p. 330). Pragmatic synonymy arises when characters “use different linguistic

devices to achieve similar ends” (p. 330), meaning that their conversational forms differ on the surface but converge toward the same goal. Pragmatic homonymy occurs when characters employ “similar linguistic devices to achieve different ends” (p. 330), producing apparent alignment at the linguistic level while pursuing divergent communicative purposes. Finally, pragmatic identity is observed when interlocutors share almost identical pragmatic strategies, creating a high degree of alignment in both form and intent. This framework sheds light on the conversational aims of participants, revealing how their individual speech styles align or conflict at the levels of form, intention, and communicative effect. Its value lies in showing how apparent conflict at the surface – such as markedly different speech styles – can mask deeper alignment at the level of communicative intent. In this paper, I draw on this framework to analyse interaction in *Skylight*, examining how linguistic strategies are structured at the levels of form, intent, and effect, and how these strategies contribute to characterisation and the dynamics of argument between the two protagonists.

In the next section, I show how looking at *Skylight* through a multimodal pragmatic lens brings out the different ways characterisation and rapport are built on stage through the mix of dialogue and other semiotic resources. This approach identifies the surface resources – gesture, space, prosody – through which interaction is materially realised; the pragmatic perspective, rooted in linguistics, accounts for how speakers manage these interactions through conversational strategies and communicative intentions. The performance I analyse here was directed by Stephen Daldry and captured live from Wyndham’s Theatre in London’s West End in 2014. I have chosen this play because it is structured entirely around the dialogue and arguments of two characters: this emphasis on verbal conflict makes it an ideal case for analysing how characterisation and interaction are constructed through both linguistic and paralinguistic means. My analysis reveals that the sense of conflict comes not only from what is said but also from the clash of contrasting speech styles and non-verbal cues. At a deeper level, however, it becomes clear that the two protagonists share more common ground than it first appears, particularly in the pragmatic strategies they use and, ultimately, in their communicative goals.

4. A multimodal pragmatic analysis of the play

Stephen Daldry’s production of *Skylight* foregrounds the conflictual, oppositional dynamics at the heart of the play, embodied in the sharply contrasting personalities of Tom and Kyra. As Daldry himself observed, the

drama is essentially “a battle between two people with two very different value systems”.⁴ Yet this confrontation is never purely combative: alongside its duelling quality, the dialogue also takes on the rhythm of a dance. The two protagonists know each other intimately, and their exchanges often carry the sense of anticipation, as if each can predict the other's moves. This interplay is constructed multimodally, through the fusion of language with paralinguistic, proxemic, and kinesic resources. Crucially, it arises from the tension between the surface features that mark Tom and Kyra as distinct individuals – differences in speech style, bodily movement, and spatial use – and the deeper layer of their communicative goals. As I will argue, while the characters diverge strikingly in how they speak and perform, their pragmatic aims are surprisingly aligned. What emerges is that the very features that seem to make their relationship impossible are, in fact, only superficial. At a more profound level, Tom and Kyra are too similar, and it is this sameness, expressed through pragmatic synonymy and pragmatic identity, that ultimately prevents them from moving forward as a couple. In order to prove my point, I shall begin by examining the communicative surface features that characterise Tom and Kyra, and I will later integrate them with a pragmatic analysis of their communicative aims.

In terms of individual style, Tom and Kyra radically differ. Tom speaks with great verbal energy, often at a rapid tempo, piling up words in a way that reflects his restless, domineering personality. His speech is characterised by verbosity and a tendency to interrupt, overlap, or pre-empt responses, which makes his dialogue feel like a performance of authority as much as an act of communication. He frequently uses rhetorical questions, exaggerations, and emphatic repetitions to assert control over the conversational space. At times Tom's tempo is so rapid that he stutters, creating the impression that his thoughts are moving too quickly for his words to keep pace. His delivery is strongly shaped by his emotional state, and one of his most striking features is his tendency to leave words unfinished, as if the words themselves were chasing one another. His preference for hyperbole lends his speech a performative, almost ostentatious quality: he often appears to be pontificating, with elaborate syntax and a tone edged with sarcasm. A particularly telling instance of his rhetorical skills occurs when Tom tells Kyra how he responded to his son's accusation that he was a cold entrepreneur:

TOM (*Spoken extremely fast*) I said perhaps it's TRUE perhaps I'm not brilliantly contemplative(.) perhaps I do not stop like some Oxford (.)

⁴ Carey Mulligan & Bill Nighy Chat About the Broadway Revival of David Hare's *Skylight*. Available at <https://www.youtube.com/watch?v=6LCD4GpR5sl>.

SMARTARSE philosopher to ask myself the “purpose of it all” but the rough effect of all my endeavour my putting my house my mortgage my car the whole of my bloody life on the line as ↑I REMINDED him I have done in my time has been to embody this unspeakably crude ASSUMPTION that it’s still worth human beings TRYING to GET SOMETHING DONE... (Act 1, 32:20-32:41)

Nighy delivers this section at high speed, with no pauses, conveying how emotionally charged Tom’s defence of his actions is. The paraverbal features – the velocity of his speech, his refusal to allow silences, the syncopated rhythm of his utterances – reflect Tom’s underlying aggressiveness. His verbal behaviour, and as I will later show, not only his verbal behaviour, resembles a siege: through his incessant tirades against banks, his complaints about young people who “don’t understand the importance of business,” and his recurring *o tempora, o mores!* stance, Tom seeks to disarm Kyra, to invade her space, and ultimately, quite explicitly, to win her back. His tone is deliberately ironic and sarcastic, coloured with a forced camaraderie after three years of absence. A clear instance of this ironic style occurs when, in response to Kyra’s request that he take off his coat, Tom sarcastically replies:

TOM Well I would of course if you’d get CENTRAL HEATING (.) Then of COURSE I’d take my coat off but since you’ve made a STYLE choice to live in Outer Siberia (.) I think (.) for the moment I’ll keep my coat on. (Act 1, 23:23-23:34)

These paralinguistic features are accompanied in Nighy’s performance by an intelligent, sapient use of proxemic and kinesic signs. The way Tom takes possession of the space is particularly relevant and indicative of both his character and his attitude toward Kyra. As previously noted, space is a crucial semiotic resource, as it can signal degrees of intimacy or social distance between characters, while its transgression may signal aggression. In this scene, Tom quite literally invades Kyra’s space: first by ringing the bell for more than sixty seconds – a strongly face-threatening act – and then by marching uninvited into her apartment. This sense of intrusion is heightened by kinesic signs, particularly Tom’s body language. From the moment he enters, he paces restlessly, handling Kyra’s books, objects, and even the cans of tomatoes she is cooking with. His incessant movement, combined with the habitual shifting of objects – at times even with his feet, as when he repeatedly nudges a chair (see figure 1) – lends Nighy’s performance a balletic grace, reinforcing the impression of a dance-like presence that both unsettles and dominates the space.



Figure 1
Tom's restlessness.

Tom's refusal to sit and his constant movement serve two distinct functions. On the one hand, they contribute to his characterisation, reinforcing the energetic personality and restlessness that Kyra later identifies as his defining quality. On the other hand, his behaviour constitutes a violation of Kyra's space, as if appropriating her and her environment; he scrutinises the external elements of her existence in order to reassert control. This dimension of Tom's character, and of his relationship with Kyra, is only briefly suggested in the script, but it becomes markedly more evident in performance, where the extent of his restlessness is made visible and coheres with his speech style.

Kyra is the absolute opposite of Tom, both from the point of view of speech style and of proxemic and kinesic signs: where Tom's speech is expansive, fast, and overflowing, Kyra's is typically more measured and restrained. She tends to choose her words carefully, often speaking in shorter, more economical turns that reflect both her self-discipline and her defensive stance in the conversations. At times, her speech takes on a didactic tone, which positions her as someone who explains, instructs, and justifies. Yet this same controlled style can also make her seem emotionally distant. Unlike Tom, who reveals his turmoil through his excessive language, Kyra often withholds, deflects, or resists, and her silences or pauses are as significant as her words. Her verbal style matches both her proxemics and kinetics: she keeps a safety distance from Tom through the entire first act, avoiding eye contact or touching. The distance she keeps from Tom during the majority of the first act shows her defensive attitude, also reinforced by her tendency to cross her arms and to stand still – thus also visually opposing Tom's constant gesturing (see figure 2 below).



Figure 2
Tom's gestures and Kyra's defensiveness.

These paralinguistic, proxemic, and kinesic features contribute significantly to Kyra's characterisation, placing her in complete opposition to Tom not only in terms of values or ideology (which are clearly conveyed by the content of their exchanges), but also by marking their differences and their conflicts at the level of behaviour, attitude, and style. It can be argued, then, that the main conflict of the play emerges in the clash of verbal and non-verbal surface features, which bring out the distance between the two characters as individuals. However, if these aspects are examined in isolation and not considered within the context of their conversational strategies and overall communicative metastructure, a stylistic interpretation of the play would remain incomplete. In fact, it is the analysis of the pragmatic competence of the two interlocutors, integrated with other semiotic resources, that provides a fuller understanding of the communicative mechanisms at work in the play, because it shows what communicative intent the different individual styles aim at.

An attentive analysis of Tom and Kyra's communicative metastrategies reveals that their interactions are characterised primarily by pragmatic synonymy and pragmatic identity (Lakoff and Tannen 1984, p. 330). This means that, in conversation, they either employ different linguistic devices to achieve similar ends, or they use similar linguistic devices directed toward the same end. At the level of communicative intent, then, Tom and Kyra are remarkably alike, while the conflict seems to operate mainly at the surface level of their exchanges, in the form of style, tone, and delivery.

A telling example of pragmatic synonymy can be found in the contrasting verbal strategies they use to project self-control and dominance over the interlocutor. Tom relies on repetitions, rhetorical questions, and an almost compulsive verbosity, strategies that allow him to assert authority and to present himself as unaffected by Kyra's decision to leave him. His manner suggests determination to control the conversational ground. Kyra, by

contrast, adopts the opposite strategies: long pauses, calculated silences, and short, clipped utterances. Yet despite the different linguistic forms, her aim mirrors Tom's: she too seeks to project indifference, to maintain an appearance of composure, and to deny Tom the satisfaction of provoking an emotional response. Both characters, in their different ways, attempt to seize control of the situation and to demonstrate self-mastery, thereby enacting pragmatic synonymy.

Many times, however, Tom and Kyra employ identical strategies in order to achieve the same communicative goals, particularly when seeking to challenge the validity of the other's perspective. These exchanges typically unfold in two stages. First, when one interlocutor presents his or her point of view, the other responds with apparent cooperation, signalling agreement through nodding or brief follow-up moves such as one-word answers. Yet once the small talk falls away and they begin to articulate their true position, they abruptly and forcefully question the validity of everything their partner has just said, thereby exposing the extent to which their initial compliance concealed their real stance. Two examples illustrate this strategy. In the first, Tom has been speaking at length about his business and his aversion to banks. As can be seen, Kyra initially appears perfectly cooperative, though not especially talkative:

TOM /.../ The banks are running the wor- you think oh fuck I'll I'll I'll run a business I'll build a business (.) You remember Kyra we started out my God it was grea- actually COUNTING the money, you counted it with me . . .

KYRA Of course.

TOM Actually HANDLING the money each morning after you'd joined us totting it up each Saturday night

KYRA I remember.

TOM Then ↑oh Christ there's this fatal moment EXPANSION

KYRA Sure. (Act 1, 28:26-28:45)

An identical situation presents itself in the second act, but this time it is Kyra who has been talking about her job and her life decisions:

KYRA Tom these are kids from VERY TOUGH backgrounds. /.../ At the very least you (.) offer them support you care for them you offer them security you give them an environment where they feel they can grow but also (.) you make BLOODY sure you CHALLENGE them (.) you make sure they realise LEARNING IS HARD. /.../

TOM I see that.

KYRA I TELL you it's fucking interesting

TOM I'm sure

KYRA FINDING that ↓balance

TOM Sure. (Act 2, 9:17- 9:59)

In both passages, the characters initially signal their understanding of what the other has said by providing very brief confirming responses (“sure”, “I remember”, “oh yeah”, “I see that”, “I’m sure”). These function as conversational strategies aimed at reassuring the interlocutor that their point is being acknowledged and, ostensibly, agreed with. However, once the exchange moves beyond small talk, they resort to almost identical phrasing to undermine and invalidate the other’s point of view. In the following examples, both employ the strategy of echoing the other’s statements to challenge their validity.

KYRA Do you think(.) do you think I’ve BELIEVED this STUFF you’ve been telling me?

TOM Stuff?

KYRA (*She imitates his voice and movements as she speaks, and speaks with no pauses*) “I’m enjoying the business, it’s wonderful I get on great with my son Alice dying was hard but of course I survived it no problem I just dropped round to see you no reason I just thought it was TIME . . .” (.) and ME (.) I’m standing here, nodding smiling agreeing like some APE and thinking (.) is this man lying to me DELIBERATELY? Or does he not even notice? Or is he so used to LYING TO HIMSELF? (Act 1 60:10- 60:22)

Tom does the exact same things a few scenes later:

TOM (*Speaking extremely fast, raising his voice*) I mean I’ve been LISTENING I’ve been listening to this STUFF you’ve been telling me ↑the BUS! The SCHOOL! even the kind of place that you choose to live and I’m thinking (.) my God my dear old friend Kyra’s joined some ↓OBSCURE RELIGIOUS ORDER (.)↑The Kensal Rise chapter! She’s performing an act of CONTRITION. (Act 2, 24: 31-24:35)

Kyra and Tom employ almost identical wording in formulating their negative evaluations of the preceding exchange (“I’m thinking, is this man lying to me deliberately?”; “I’m thinking, my dear old friend Kyra’s joined some obscure religious order”). Both resort to superficial communicative strategies that signal apparent agreement, only to reveal that they have concealed their true thoughts and, in fact, not only disagree with but also despise each other’s point of view. Pragmatic identity is revealed here in the way the protagonists deploy the same tactics to achieve the same goal: persuading the other of the absurdity of their respective positions.

Pragmatic synonymy and pragmatic identity reveal the profound nature of the interactional and communicative structure of these two characters, which is based on non-acceptance, and ultimately on asserting control and power over the other person. It is at this level that the conflict of power really emerges and becomes unsolvable, in that “in the inter-personal sphere, attempted changes in beliefs can also function as assaults on the mind of one

participant by another and as a means to power” (Herman 2005, p. 236). This analysis suggests that what renders the relationship impossible – and the conflict so powerful on stage – is not merely the superficial incompatibility of values, but rather the absolute identity of communicative intent. Both characters are driven less by a desire to understand one another than by the determination to “win the argument”, and it is this dynamic that ultimately prevents them from moving forward together. On the one hand, Kyra wants Tom to admit that what he’s been doing for his entire life is soulless, arid, oppressive:

KYRA I mean, like earlier (.) earlier this evening. you were going on about BUSINESS (.) no one understands business that’s what you said (.) Suddenly I must say I hear it everywhere. (*Talking very slowly*) These so-called ACHIEVERS telling us they have a grievance (.) The WHOLE OF SOCIETY must get down on their KNEES and thank them, because they do something they no longer call MAKING MONEY. Now we must call it something much NICER (.) Now we must call it ‘the creation of wealth’. (Act 2, 33:34- 33:50)

Tom, on his part, wants Kyra to admit that she is wasting her talent, that her choice is futile, oppositional, immature:

TOM Kyra, ↑ look at you now! it won’t even work it can’t work (.) because it’s built on a negative (*he stutters*) it’s it’s it’s built on escape (.)WHAT is it in you? This thing that you have? Why doesn’t it YIELD? ↓I don’t UNDERSTAND it. (Act 2, 38:59-39:11)

Unable to accept the other for who he or she is, Tom and Kyra are imprisoned in their own trenches (see figure three below). It is this shared rigidity, conveyed through similar pragmatic strategies, that prevents the achievement of compromise or common ground, and ultimately results in non-communication at the deepest level, “that of the effect on the other participant” (Lakoff and Tannen 1984, p. 344). The analysis of these three levels of communication, thus, helps to explain both the battle and the dance that characterise the dialogue of the play: the battle is most visible at the level of surface features, where Kyra and Tom appear radically different, while the dance emerges at the deeper levels, where the similarity of their pragmatic strategies and communicative intentions becomes apparent.

To conclude, it can be argued that *Skylight* cannot be fully explained either by differences in ideology or by contrasts in style alone. The semiotics of theatre makes visible how meaning is embodied in gesture, space, and paralinguistic features, while pragmatics reveals the deeper identity of communicative intent that drives the exchanges. Taken together, these approaches show that Tom and Kyra’s relationship collapses not simply because of their ideological distance, but because of their shared

communicative rigidity: they are locked in a struggle where both seek victory rather than mutual understanding.



Figure 3
The position of Tom and Kyra on stage.

5. Conclusions

This paper has argued that a multimodal pragmatic perspective can account more fully for how *Skylight* communicates meaning in performance than approaches that privilege either the written text or performance in isolation. By examining paralinguistic, kinesic, and proxemic resources alongside dialogue, the analysis has shown that what appears to be a stark conflict between Tom and Kyra at the level of surface features – speech tempo, rhythm, gesture, spatial use – coexists with a deeper alignment in communicative intent. The concepts of pragmatic synonymy and pragmatic identity help to articulate this paradox: the protagonists often pursue similar interpersonal aims while enacting them through contrasting verbal and non-verbal strategies. This layered patterning explains why the play reads as both battle and dance, opposition and attunement, rupture and recognition.

The findings also underscore a methodological point. Treating recorded theatre as embodied discourse permits close, revisitable attention to interactional details that typically elude page-bound readings. In the 2014 production examined here, the orchestration of language with movement and space makes the complexity of Tom and Kyra's relationship legible: Nighy's high-velocity prosody and insistent spatial encroachment meet Mulligan's controlled economy of turn-taking and guarded proxemics, producing a communicative texture in which similarity of aim is masked by dissimilarity of form. Multimodal pragmatics provides the descriptive tools to track that texture and to show how characterisation, rapport, and power are negotiated

in real time on stage.

At the same time, the analysis remains provisional in two respects. First, it focuses on a single production, and therefore on one legitimate but non-exhaustive realisation of the play's semiotic potential. Second, it draws on a filmed performance, which, while valuable for systematic observation, is itself a mediated interpretation. Future research might test the robustness of these claims by comparing multiple stagings and performance styles, or by extending the framework to different dramatic genres and ensemble dynamics where multimodal coordination is distributed across larger casts.

More broadly, the approach set out here has implications for stylistics and theatre studies. For stylistics, it suggests that the interpersonal meanings of dramatic dialogue are best understood when verbal choices are analysed together with the embodied choices that accompany them. For theatre studies, it offers a principled way to describe how ideological tensions are dramatised not only in what characters say but in how they time, place, and embody saying it. In *Skylight*, the ultimate dramatic insight is that communicative failure does not stem from irreconcilable differences alone; it also arises when two speakers are, beneath the surface, too alike in their pragmatic aims. A multimodal pragmatic lens makes that insight analytically visible.

Bionote: Valentina Vetri teaches English Language and Translation at the University of Siena. She holds a PhD in Translation Studies from the University of Birmingham, UK. As a translator, she has collaborated with the publishing houses Rusconi, Rubbettino, and Cue Press. She has published academic articles on linguistics, translation and pragmatics. Her latest monograph, *Poetics, Ideology and Dissent: Beppe Fenoglio and Translation*, was published by Palgrave MacMillan in 2023. She is currently working on the translation of the critically acclaimed play *Hansard*, written by Simon Woods. She is part of the international research network "Coercive Control: From Literature into Law", led by the Universities of Leeds and York, where she investigates the representation of linguistic manipulation and gaslighting within domestic abuse. Her research interests include pop culture linguistics, stylistics, and translation studies.

Author's address: valentina.vetri@uni.si.it

References

- Busse B. 2011, *Teaching the Stylistics of Drama*, in Jeffries L. and McIntyre D. (ed.), *Teaching Stylistics*, Palgrave Macmillan, London, pp. 152-177.
- Culpeper J. 2002, *(Im)politeness in Dramatic Dialogue*, in Culpeper J., Short M. and Verdonk P. (ed.), *Exploring the Language of Drama: From Text to Context*, Routledge, London, pp. 83-95.
- Culpeper J. 2014, *Language and Characterisation: People in Plays and Other Texts*, Routledge, London.
- Elam K. 2002, *The Semiotics of Theatre and Drama*, Routledge, London.
- Fischer-Lichte E. 1992, *The Semiotics of Theatre*, Indiana University Press, Bloomington (IN).
- Furlong A. 2014, *Outsourcing: A Relevance-Theoretic Account of the Interpretation of Theatrical Texts*, in Chapman S. and Clark B. (ed.), *Pragmatic Literary Stylistics*, Palgrave Macmillan, London, pp. 70-89.
- Ghazvinieh A. 2025, *Performance as Discourse: A Multimodal Perspective on Meaning-Making in Theater*, in "Semiotica" 266, pp. 171-191.
- Halliday M.A.K. 1985, *An Introduction to Functional Grammar*, Edward Arnold, London.
- Halliday M.A.K. 1993, *Language in a Changing World*, Applied Linguistics Association of Australia, Deakin.
- Hare D. 2008, *Skylight*, in *Plays Three*, Faber & Faber, London, pp. 7-110.
- Herman V. 1991, *Dramatic Dialogue and the Systematics of Turn-Taking*, in "Semiotica" 83 [1-2], pp. 97-121.
- Herman V. 2005, *Dramatic Discourse: Dialogue as Interaction in Plays*, Routledge, London.
- Kress G. and van Leeuwen T. 1996, *Reading Images: The Grammar of Visual Design*, Routledge, London.
- Lacey S. 1999, *Public Spaces and Private Narratives: The Plays and Films of David Hare*, in Bignell J. (ed.), *Writing and Cinema*, Routledge, London, pp. 150-163.
- Lavender A. 2021, *Multimodal Acting and Performing*, in Elleström, L. (ed.), *Beyond Media Borders, Volume 1: Intermedial Relations among Multimodal Media*, Palgrave Macmillan, Cham, pp. 113-140.
- Lakoff R. and Tannen D. 1984, *Conversational Strategy and Metastrategy in a Pragmatic Theory: The Example of Scenes from a Marriage*, in "Semiotica" 49 [3-4], pp. 323-346. <https://doi.org/10.1515/semi.1984.49.3-4.323>.
- Macrae A. 2014, *Stylistics, Drama and Performance*, in Burke M. (ed.), *The Routledge Handbook of Stylistics*, Routledge, London, pp. 253-267.
- Martin C.A. 1997, *Staging the Reality Principle: Systemic-Functional Linguistics and the Context of Theatre*, Macquarie University, Thesis, <https://doi.org/10.25949/19433216.v1>.
- McAuley G. 1994, *The Video Documentation of Theatrical Performance*, in "New Theatre Quarterly" 10, pp. 183-194. <https://doi.org/10.1017/S0266464X00000348>.
- McIntyre D. 2008, *Integrating Multimodal Analysis and the Stylistics of Drama: A Multimodal Perspective on Ian McKellen's Richard III*, in "Language and Literature" 17 [4], pp. 309-334.
- National Theatre Live: *Skylight*, dir. Stephen Daldry, National Theatre of Great Britain Production, 2015.
- O'Halloran K.L., Tan S. and K.L.E. M., 2014, *Multimodal Pragmatics*, in Schneider K.P. and Barron A. (ed.), *Pragmatics of Discourse*, De Gruyter Mouton, Berlin, pp. 239-

268.

- Short M. 1998, *From Dramatic Text to Dramatic Performance*, in Culpeper J., Short M. and Verdonk P. (ed.), *Exploring the Language of Drama: From Text to Context*, Routledge, London, pp. 6-18.
- Tan S., Wignell P. and O'Halloran K.L. 2016, *Multimodal Semiotics of Theatrical Performances*, in Sindoni M.G., Wildfeuer J. and O'Halloran K.L. (ed.), *Mapping Multimodal Performance Studies*, Routledge, London, pp. 26-50.
- Wu D. 2007, *Hare's Stage Poetry, 1995-2002*, in Boon R. (ed.), *The Cambridge Companion to David Hare*, Cambridge University Press, Cambridge, pp. 79-91.

THE MULTIMODAL EXPRESSION OF GRATITUDE

The Interplay of Words and Pictures in Thank-you Cards¹

ELISA BERTOLDI, SARA GESUATO
UNIVERSITY OF PADUA

Abstract – This paper examines the linguistic and visual components of over 90 print, non-pop-up thank-you cards within the frameworks of speech act set analysis (Blum-Kulka *et al.* 1989) for the linguistic component, and Kress and van Leeuwen’s (2021) social semiotic analysis, integrated with Unsworth and Cleirigh’s (2017) framework on intersemiotic relations, for the visual component. The linguistic component of the texts is analysed in terms of head acts and supportive moves, identified partly inductively and partly based on the literature (Ahar and Eslami-Rasekh 2011; García and Kleifgen 2016; Eisenstein and Bodman 1986, 1993; Schauer and Adolphs 2006). The visual component is examined through the analysis of representational structures and compositional arrangements, paying special attention to the intersemiotic relations between verbal and visual elements. The cards include direct, occasionally repeated, head acts (e.g., “thank you”; “With sincere gratitude and warmest thanks”). These are often typographically enhanced through bolding or capitalisation and are sometimes characterised by plays on words. The cards may also include supportive moves, such as complimenting the benefactor (“life saver”) and acknowledging the impact of the benefit (“That means everything”). The analysis reveals three primary intersemiotic relationship types: humorous engagement, whereby visual elements enable or reinforce intersemiotic wordplay; aesthetic enhancement, whereby images provide visual pleasure; and semantic support, whereby visuals clarify or emphasise verbal expressions. Food imagery most often serves a humorous function, while floral elements primarily provide aesthetic enhancement, and animal imagery demonstrates versatility across all relationship types. Overall, the study demonstrates that multimodal thank-you cards employ intersemiotic coordination strategies that transform routine politeness acts into opportunities for social bonding, suggesting that multimodal analysis can enhance our understanding of how different semiotic resources are co-deployed in positive facework.

Keywords: thank-you cards; multimodality; intersemiotic relations; verbal text; visual text.

¹ Both authors designed the study and analysed the data. They jointly wrote Sections 2.3., 3.1, 3.2, 3.3, 3.3.4, 5.1 and 5.2. Elisa Bertoldi also wrote Sections 2.2, 3.3.2, 3.3.3, 4.1.3, 4.2 and 4.3. Sara Gesuato also wrote Sections 1, 2.1, 3.3.1, 4.1.1 and 4.1.2.

1. Introduction

Pragmatics sees language, or rather language use, as a form of goal-directed social behaviour. It thus explores how people use language to perform actions and enact social relations in context. As part of purposeful communicative behaviour, speech acts (i.e., verbal actions) are privileged means for achieving goals and regulating interpersonal relationships. Expressing gratitude is one such purpose- and relation-oriented verbal action: it both expresses a psychological state and performs the interactional function of balancing out social debts and credits.

Prototypically, thanking is the verbalisation of the feeling of gratitude and indebtedness that a speaker/writer feels toward the addressee as a result of something beneficial, or at least intended as beneficial, that the addressee intentionally performed for the speaker/writer or provided him/her with. It is meant as an acknowledgement of the addressee's generous behaviour and reciprocation for it.

Thanking is a harmony-preserving interactional contribution which helps regulate interpersonal relationships. First, it indicates that the thanker is well-disposed towards the thankee; indeed, as an expressive speech act (Searle 1976), it conveys the thanker's psychological state of gratitude toward the thankee for a past action that benefited the thanker. Second, it enhances the thankee's positive face by acknowledging their previous laudable deed, thus making them feel good; that is, it functions as a positive politeness strategy (Brown and Levinson 1978). It also preserves the thanker's positive face, because their courteous behaviour makes them appear likeable. Third, in its very acknowledgement of the thanker's state of dependence on the addressee, it functions as a verbal repayment of the debt incurred in accepting the thankee's benefit. Thus, if an act of thanking is successful (i.e., if it is understood, accepted, and possibly ratified with a relevant response), the interactants' social debts and credits will have been balanced out.²

An act of thanking may be conventionally routinised or originally elaborate, depending on several variables. These include the actual or perceived magnitude of the benefit received, and its long-term effects; the role-relationship between the interactants (their degree of familiarity and social (a)symmetry); the context in which the interaction takes place, which

² However, an act of thanking is not a failsafe interactional mechanism. It threatens the thanker's negative face, since indebtedness involves an obligation to return the favour received. Also, if it is not accepted (e.g., because it is not considered adequate or because the thankee wants to break off their relationship with the thanker), the thanker's positive and negative face will be damaged: on the one hand, he/she will feel rejected, disliked, unappreciated; on the other, he/she will remain indebted.

can be private or public, informal or formal; the medium (e.g., written) and instrument (e.g., the phone) of communication, which can make the interaction synchronous or asynchronous, and ephemeral or permanent; the personality of the thanker, who can be effusive or reserved; and the conventional cultural expectations holding in the context where thanking takes place, which shape interactants' choice of content and wording.³

The asynchronous and monologic nature of written thanks brings about two main effects.⁴ On the one hand, it removes the pressure of immediate response management, which allows writers to plan, ponder, and revise their texts before they are delivered. On the other hand, it produces potentially permanent texts, which can be stored, shared and accessed multiple times across space and time. Written thanks, therefore, both allow and call for greater elaboration and more complex face-work strategies than their synchronous spoken counterparts. This renders them suitable for formal and official occasions where they can evolve into established genre conventions (e.g., book acknowledgements).⁵ One established macro genre where written thanking occurs is that of greeting cards. These are commercially produced, multimodal texts that blend verbal and visual elements, and count as culturally ritualized vehicles for expressing, among other things, gratitude.

Commercially printed greeting cards are a macro genre in which the roles of Principal, Author and Animator (Goffman 1981) shift between the card buyer/sender and the card copywriter/designer/manufacture. At the planning/design stage, the Animator is the manufacturer, who physically produces the card; the Author is the copywriter and/or designer, who select(s) and/or create(s) verbal and visual content; the Principal is initially the brand, which chooses what stance to project (e.g., funny, sentimental, elegant). At the delivery stage, the Principal role shifts to the card sender, who appropriates the message as an expression of their own stance. That is, a ready-made resource is crafted by a professional team for potential use in a number of contexts; when it is selected, delivered and possibly personalised by a sender, it is recontextualised for a specific communicative occasion.

Greeting cards are comparable to other multimodal genres which combine words and still pictures: picture books (Moya-Guijarro 2016), traffic signs (Forceville and Kjeldsen 2018), memes (Vásquez and Aslan 2021), magazine covers (Tseronis 2015), posts (Bouko 2020), public health posters (Al-Subhi 2024) and tweets (Oakey *et al.* 2022), toiletries ads (Del Saz-Rubio

³ Determinants of gratitude expressions may also include subtle variables like eye contact between the benefactor and the beneficiary (Okamoto and Robinson 1997).

⁴ This is intended to the exclusion of texting, which is fingered speech, not writing (McWhorter 2012).

⁵ Given the ease of recording speech, spoken thanks can also become permanent in both public contexts (e.g., at awards ceremonies) and private ones (e.g., at birthday parties).

2018), and postcards (Francesconi 2011). Yet, linguistic-multimodal analyses of greeting cards are scarce (Atminytė 2022). In particular, to our knowledge, thank-you cards have not been systematically explored in linguistics, particularly regarding how their multimodal nature contributes to the performance of thanking speech acts.

In the remainder of the paper, we provide a brief overview of the literature on the act of thanking and on the analysis of multimodal texts (Section 2), and then we illustrate our research method (Section 3). Next, we report our findings (Section 4). Finally, we discuss and draw implications from them (Section 5).

2. Literature review

2.1. *Thanking*

Studies on the speech act of thanking have identified its two core structural components: head acts and supportive moves. The head act is the gist of thanking, which minimally and explicitly conveys the speaker's/writer's illocution through conventionally recognised plain or emphatic formulae (e.g., “Thank you”, “Many thanks”, “I owe you one”, “Much appreciated”), and which may also specify the object of the illocution (e.g., “Thank you for having me”, “I really appreciate what you did for us”). A supportive move is an elaboration on the head act that is meant to enhance the effectiveness and appropriateness of the gratitude being expressed. It refers to the situational factors that give rise to, and thus motivate, the sender's gratitude, such as the positive tangible effects of the benefit (e.g., negative consequences avoided), the positive emotional impact of the benefit on the beneficiary (e.g., relief, joy), the benefactor's effort and dedication (e.g., commitment, generosity), or the actual or perceived value of the benefit (e.g., its cost, its durability, scope).

Scholars have reported the variable realisation of acts of thanking, proposing partly different taxonomies for analysing thanking strategies.

Coulmas (1981) classified speech acts of thanking along multiple categories: as *ex-ante* (i.e., for a potential good or service) vs. *ex-post* (i.e., for a good or service already provided); as relevant to material vs. immaterial goods, as originating from requested vs. non-requested benefits; and as involving indebtedness vs. non-indebtedness. Instead, Eisenstein and Bodman (1986, 1993), who examined role-play data, classified thanks into three main types: phatic, ritualised responses; expressions of thanks accompanied by subsidiary acts like jokes or questions; and speech act sets, namely expressions of gratitude qualified by additional acts like complimenting, reassuring, or promising to repay.

In DCT-elicited data, Held (1996) identified supportive moves, such as compliments to benefactors and acts of self-denigration, as well as remedial moves that emphasise the thanker's indebtedness and/or the intention to reciprocate. Similarly, in their study of DCT-elicited and corpus data, Schauer and Adolphs (2006) identified numerous supportive moves of gratitude expressions, namely complimenting the interlocutor, stating the reason, confirming the interlocutor's commitment, stating the intent to reciprocate, stating the interlocutor's non-existent obligation, and refusing.

Díaz Pérez (2005) produced a taxonomy that identifies thanking strategies differing in their register appropriateness (formal, e.g., *I give you a lot of thanks*; unmarked, e.g., *Thank you* and colloquial like *Cheers* or *Ta*), and specific supportive moves that may accompany or reinforce them. These include praise to the action (recognising the benefit received) and praise to the person (complimenting the benefactor directly). By contrast, Cheng's (2005) taxonomy comprises eight strategies: thanking, appreciation, positive feelings, apology, recognition of imposition, repayment, other expressions, and alerters.

Although most pragmatic research on thanking has focused on spoken interaction, some studies have examined written thanks such as acknowledgements in publications. The pioneering study by Ben-Ari (1987) showed that their content, structure and formulation serve to construct and manage socio-professional meanings. For example, it highlighted complex dynamics of debts and obligations in the academic community, often masking a hostile tension between respect for hierarchies and a need for independence. It also presented writers as individuals in a network of supportive interpersonal relationships.

Cronin (1995) identified the various dimensions along which acknowledgements may vary: the richness of their content, the intensity of the feelings expressed, the number of acknowledgees mentioned, the goods they are relevant to, the writers' private goals (e.g., cementing relationships).

Giannoni (1998, 2002) examined journal article acknowledgements, showing that they contain optional opening and closing moves (i.e., establishing ancestry, anticipating future interest in the article) and a compulsory credit-giving move, each expandable in specific sub-components. They were presented as manifestations of the authors' civility and membership in the scientific community, which are intended to acknowledge dependence on others for the success achieved in a competitive environment. Giannoni's (2006) analysis of book acknowledgements focused on features relevant to the texts' evaluative function, namely hyperbole, irony and emotivity, which were found to be prevalent in the soft sciences. They appeared to serve a politeness function: maximising praise of the acknowledgees, while downgrading oneself. The goal of such texts was

reported to be the furthering of the writers' careers by making explicit their network of relationships.

Hyland (2003, 2004) and Hyland and Tse (2004) studied acknowledgements in graduate dissertations, coding them for benefit categories (mostly time and academic assistance, but also technical, financial and moral support) and describing them for the sequencing of the benefactors mentioned in them (typically: academic, peers, and friends and family, in this order). The authors described their structure as consisting of three main moves: an optional Reflecting Move (i.e., a comment on the writer's research experience), an obligatory Thanking Move (i.e., mapping credit to individuals and institutions), and an optional Announcing Move (i.e., a statement of responsibility and inspiration).⁶ The formulation of the acknowledgement moves was found to be restricted to a few lexico-grammatical patterns. The authors concluded that this genre allows writers to project both personal and public identities, and to promote their professional identity as competent and well-connected members of a discourse community, binding acknowledgers and their academic acknowledgees in a mechanism of long-term mutual indebtedness.

Studies relevant to thanks in general, and to written public thanks, in particular, reveal the range of strategies used to manage social relationships and negotiate face-work. Indeed, gratitude is shaped by social and situational factors, such as status and the level of imposition, as well as emotional and evaluative stance. Since each act of thanking differs from others in terms of sincerity, level of indebtedness and context, it can be variously expressed, involving multiple components, modifiers, and supportive elements.

Studies more specifically focused on acts of thanking found in publication acknowledgements involve managing conflicting social needs (face wants), and raise questions about the interactional purposes they serve, the perspectives they reflect, and the underlying messages they convey. Although they typically follow a structured sequence, they differ in content and style across languages, disciplines and individuals.

The literature reviewed above shows that rich frameworks have been put forward for analysing verbal thanking strategies and their accompanying facework, which can be usefully applied also to the analysis of thank-you cards.

⁶ The structure reported by Hyland (2004) was also documented in Al-Ali's (2006) study of PhD dissertation acknowledgements written in English by Arab speakers, but with some variation in their component steps.

2.2. Multimodal texts

Research across multiple multimodal genres has demonstrated that successful multimodal texts achieve effects that are far greater than the sum of their parts through the deployment of intersemiotic relationships (e.g., Liu and O'Halloran 2009). These operate along several key dimensions.

For example, Liu (2023) showed how multimodal texts employ both convergent and divergent relationships between modes, creating elaborate communicative effects. Similarly, Yu's (2019) analysis demonstrated how visual and verbal elements can serve distinct communicative functions simultaneously, creating interpretive spaces that allow for complex meaning negotiation.

Unsworth and Cléirigh's (2017) research demonstrated that visual and verbal elements can provide complementary information units that enhance or complete one another. Thus, visual elements can provide spatial, qualitative, and compositional information, while verbal elements can provide categorical, sequential, and abstract information. Their synergistic relationship leads to the construction of meanings that neither mode could achieve independently. Furthermore, the authors' examination of multimodal reading processes demonstrated how meaning construction in image-text combinations requires readers to navigate complex interpretive processes that integrate information across multiple semiotic channels.

Extensive research across multimodal artefacts featuring brief text-image combinations has confirmed the broad applicability of key theoretical principles regarding visual-verbal coordination. In picture book studies, researchers have identified sophisticated strategies for integrating visual and verbal storytelling, with distinct modal relationships serving various narrative functions (Moya-Guijarro 2016; Nikolajeva and Scott 2006; Sipe 1998). Advertising research has similarly developed frameworks to explain how visual and verbal elements collaboratively produce persuasive effects that exceed the sum of their parts (Hidalgo-Downing and Kraljević Mujic 2015; Stöckl 2024), highlighting the interdependent nature of modes, whose roles shift across contexts (Guo and Feng 2017). In comics research, scholars have demonstrated how visual-verbal integration ranges from strict separation to full fusion, enabling both conventional representation and creative expression (Cohn *et al.* 2017; Skwarzyński 2019). Lambeens and Pint (2015) proposed a framework capturing how comics combine recognisable conventions with innovative elements that provoke engagement through sensory and emotional impact.

A study on contemporary postcards revealed how these multimodal artefacts function through coordination of written messages and images to create coherent communicative experiences that enable users to accomplish complex emotional work (Francesconi 2011). In her research on Edwardian

greeting postcards, Gillen (2013) analysed how pre-packaged visual vocabulary allowed senders to communicate intensely emotional arguments for maintaining their relationships without ever having to put these into their own words.

Hiippala's (2007) comprehensive multimodal analysis of tourist brochures provided particularly valuable insights for brief text artefact research. His study demonstrated how texts operating within a limited semiotic space require careful design when placing and planning text and images to fulfil their communicative purpose, while also adhering to aesthetic principles. Using systemic functional analysis, Hiippala showed how such artefacts imply the orchestration of multiple communicative functions (i.e., promotional, informational, and aesthetic) within the constraints of the semiotic space.

Finally, Forceville and Kjeldsen (2018) examined unusual and creative traffic signs, which, despite their coded nature, can be used as a form of visual "speech acts" to convey novel meanings. The authors demonstrated how their meaning is conveyed through verbal language and through the interaction of visual elements, genre conventions, and situational awareness.

This body of research establishes several key principles for analysing multimodal thank-you cards. First, cards can be understood as texts where meaning emerges from the coordination of visual and verbal elements. Second, the effectiveness of thank-you cards depends on their ability to create meanings that exceed what either visual or verbal elements could accomplish independently. This requires a deep understanding of how different modal relationships can be employed to achieve multiple communicative goals simultaneously.

2.3. Implications

As the above literature review demonstrates, numerous investigations have been conducted on verbal acts of thanking and various multimodal texts. However, no study has investigated how thanks are realised when expressed both verbally and visually. The present study examines how the speech act of thanking is realised multimodally in thank-you cards, private texts chosen by an addresser, but written by others, intended for exchange in private or semi-private settings (i.e., informal social gatherings). The research aims to identify the structural components of thank-you cards, to describe their verbal and visual realisation, to determine the types of relationships holding between their verbal and visual components, and to document the degree of variation in their realisation patterns. To this end, it develops and applies an analytical approach for examining intersemiotic relations in thank-you cards. By incorporating insights from speech act theory, visual semiotics and multimodal discourse analysis, it proposes a comprehensive framework for

understanding how cards accomplish their communicative work.

3. Method

3.1. Research questions

We addressed the following questions:

- RQ-1: What are the verbal head acts and verbal supportive moves that constitute the multimodal speech act of thanking in thank-you cards?
- RQ-2: How often and how widely are they used in multimodal thank-you cards?
- RQ-3: What typographic devices characterise the verbal speech act of thanking?
- RQ-4: What informational-functional role do visual components have in multimodal thank-you cards?
- RQ-5: How often and how widely are they used in multimodal thank-you cards?
- RQ-6: What kind(s) of relationship(s) is/are attested between the verbal and visual components of multimodal thank-you cards, and how often?

The questions were operationalised as follows:

- RQ-1-a: Which of the prototypical expressions of gratitude are attested in the multimodal thank-you cards considered? RQ-1-b: Which of the main supportive moves of acts of thanking described in the literature are attested in those cards?
- RQ-2: What are the frequency and dispersion values of the head acts and supportive moves identified in the multimodal thank-you cards considered?
- RQ-3-a: What different typographic resources (e.g., capitalisation, boldface, italicisation) are used in the verbal component of the speech acts of thanking? RQ-3-b: How often are they used?
- RQ-4: Of the informational-functional roles that visual components may play in multimodal texts, as reported in the literature, which ones are attested in the multimodal thank-you cards considered?
- RQ-5: What are the frequency and dispersion values of the informational-functional roles of visual components in multimodal thank-you cards?
- RQ-6-a: Of the relationship(s) that may hold between the verbal and visual components of multimodal texts, as reported in the literature, which

ones are attested in the multimodal thank-you cards considered? RQ-6-b: What are their frequency values?

3.2. Data collection

The corpus for this study consists of 97 thank-you cards. Of these, 17 thank-you cards were collected by one of the authors and a friend over time, simply out of personal interest in greeting cards. The remaining 80 thank-you cards were collected from *Etsy*, a global digital marketplace specialising in handmade and creative goods. The reason for this choice was that *Etsy* provides access to card designs that reflect current multimodal communication practices and that are produced by a diverse range of independent designers working across different cultural contexts.

Data collection on *Etsy* was conducted in early 2025, using four filters to restrict the search to cards that met specific criteria. This was intended to ensure corpus homogeneity. Cards had to display both textual and visual elements within the same spread, allowing us to examine semantic relationships between the two semiotic modes of communication. Additionally, cards had to be priced at or below 20 euro to represent accessible, mainstream designs rather than luxury or highly specialised products. Furthermore, cards had to be offered by “Star Sellers” only, a designation indicating established sellers with proven track records for quality, customer satisfaction, and reliable service; this was meant to ensure a baseline level of design competence and commercial viability. Finally, cards had to have their verbal component encoded in English, which would ensure language consistency across our data; however, by indicating a language that is widely used also by non-native speakers, cards would be collected from sellers across different geographic regions, thus possibly exemplifying the use of English as an international language. The search returned many results. From these, the first 100 cards were pre-selected, each offered by a different seller. The goal was to obtain a representative coverage of their heterogeneity, while avoiding individual designer bias. A few accidental duplicates were then discarded. The final selection consisted of 80 cards.

All 97 cards were later converted into digital format as .jpg files and uploaded to the *Padlet* platform to ensure shared access for both authors.

3.3. Data analysis

The analysis of the cards focused first on their linguistic and visual components, and then on the intersemiotic relationships between these two components. The approach integrated speech act analysis with social semiotic analysis to capture how the verbal and visual elements work together to accomplish the pragmatic functions of thanking.

Each card was assigned a numerical identifier, and both the numerical value and the verbal component of the card were entered into an *Excel* spreadsheet. Each card was then coded across multiple dimensions: linguistic features (head act types, supportive move categories, exact text transcription, typography; see Section 3.3.1), visual characteristics (content categories, functional roles, compositional features; see Section 3.3.2), intersemiotic relationships (integration levels, relationship types; see Section 3.3.3). More specifically, we applied an annotation scheme that enabled us to classify the verbal and visual components of each card, as well as the relationships between these components.⁷

3.3.1. Linguistic analysis

The verbal component of each card was linguistically analysed following the speech act set analysis approach developed by Blum-Kulka *et al.* (1989), which involved the identification of head acts (i.e., core thanking expressions) and supportive moves (i.e., ancillary speech acts favouring the successful performance of the act of thanking).

We considered core thanking expressions to be only those containing words like *thank you*, *thanks*, *thankful*, *grateful*, *gratitude*, and *appreciate*, which encode positive emotional reactions (coded as EXP for ‘Explicit thanking expression’). When no such explicit, prototypical gratitude expression was present, we did not “upgrade” a supportive move to head act status, as is done in Blum-Kulka *et al.* (1989). Instead, we marked the relevant verbal speech act as lacking a head act (coded as ITE for ‘Implied thanking expression’). For example, we decided that expressions like *I am indebted to you*, if attested, would be classified as realisations of supportive moves. For each head act, we also specified if it was internally modified by an intensifier (INT; e.g., *so much*) or mitigator (MIT; e.g., *a little note to say*), modulating its force; if it mentioned the benefit (BIT, i.e., the object of the illocution) and/or the benefactor (BEN), these being elements previously attested in analyses of written thanks (e.g., Hyland and Tse 2004; Al-Ali 2006); if it was characterised by humorous wordplay (HUM, e.g., *fanks for thanks*); and if it was repeated (DOB), thus giving rise to so-called *gratitude clusters* (Schauer and Adolphs 2006, p. 126).

We also coded the verbal speech acts for the presence (Y for ‘Yes’) or absence (N for ‘No’) of supportive moves. Then, drawing on the taxonomies by Cheng (2005) and Eisenstein and Bodman’s (1986, 1993), we determined if the verbal component of the cards included these possible supportive moves: Acknowledging the imposition (ACK, e.g., *I know you’re busy*),

⁷ The coding scheme is available from the authors upon request.

Complimenting the addressee (PLI, e.g., *You're amazing*), Expressing indebtedness (DEB, e.g., *I owe you*), Stating the impact of the benefit (IMP, e.g., *It means a lot to me*) or Promising reciprocation (REP, e.g., *I'll make it up to you!*). The identification of the supportive moves was not based on a predetermined list of possible lexical realisations, but rather on our joint interpretation of the propositional content of the verbal speech acts. This inductive approach led to the identification of one type of supportive move, namely the quotation of a proverb or saying, or the statement of a general reflection (PRO, e.g., *It takes a village*). Additionally, each supportive move was coded for the possible presence of lexical upgraders, specifically Intensifiers (INT, e.g., *very much*) and Wordplay (PUN, e.g., *You're GRAPE*).

When a given verbal component or feature in a card was compatible with more than one code within a coding categories, multiple codes were specified, separated by + symbols, in order to capture the richness of these multimodal texts. For example, the head act of Card 3, namely, *Thanks for being my unpaid therapist* was coded as EXP+INT, signalling that it had an explicit expression of gratitude and an intensifier. Similarly, the supportive moves of Card 21, namely *You're otterly amazing*, *You're # 1* were coded as PLI+PLI, signalling the presence of two compliments to the benefactor.

Finally, the entire verbal speech act was coded for its typographic features, that is, the presence, on all or some of its words, of Standard text (STD), Boldface (BOL), All caps (CAP), Italics (ITL), Colour-coding of morphemes/words (COL), Glitter (GLI), Embossing (EBS) and/or Incision (INC).

Here follows an example of the complete classification of the verbal component of Card 78:

Head Act Coding:

HA_Type: EXP

HA_Text: "Thank you"

Supportive Moves Coding:

SM_Present: Y

SM_Types: PLI

SM_Text: "You're indispensable"

Typography: CAP+COL

3.3.2. Visual analysis

Visual analysis drew upon Kress and van Leeuwen's (2021) social semiotic framework and was adapted to account for representational structures (what

is depicted) and compositional arrangements (layout, prominence, spatial relationships).

For each card, a verbal description of the visual content was provided in the *Excel* file. Then, the main visual elements on the front cover (and the inside spread, when present) were classified into these types: flowers/plants (FLO), animals (ANI), food/drink (FOO), abstract/geometric patterns (ABS), typography-only designs (TYP), scenes/landscapes (SCE), objects (OBJ), and people (PEO). These labels could also be used in combinations, if applicable. For example, Card 78 was classified as OBJ for its depiction of colourful medicine bottles with pills spilling out; instead, Card 40 was classified as ABS+FLO+ANI+OBJ, since it displays heart shapes and circular patterns with butterfly and decorative swirl elements in pink and purple tones, where the swirl elements appear to be plant-like botanical features.

The visual content was also classified along these dimensions: the distinction between narrative (NAR) and conceptual (CON) representations differentiated dynamic processes (i.e., actions or transformations; e.g., *three anthropomorphised dancing pastries*, Card 1) from static displays (i.e., static situations; e.g., *watercolour floral border pattern*, Card 37). The realistic (REA) versus stylised (STY) distinction refers to the degree of photographic realism versus cartoon/simplified rendering of the pictures. Compositional positioning analysed spatial arrangement, distinguishing centrally positioned elements (CEN), which indicate high visual prominence, from marginally positioned elements (MAR), which serve decorative functions. Framing was differentiated between compositions with strong borders/frames (FRA) that clearly delineate visual spaces and open compositions (OPE) featuring flowing, unbounded elements without clear boundaries.

For cards with both front and inside visual elements, we examined how visual themes were maintained or developed across the card's interior spread. We coded interior elements by applying the same scheme, thus capturing visual continuity or contrast between front and inside presentations.

Here follows an example of the complete classification of the visual component of Card 78:

VT_Description: Colourful medicine bottles with pills spilling out

VT_F_Element(s): OBJ

VT_Content and style: CON+STY+CEN+FRA

VT_I_Element(s): – (not applicable for FRONT distribution)

3.3.3. *Intersemiotic analysis*

The interplay of verbal and visual modes was analysed by drawing on the theoretical framework established by Unsworth and Cléirigh (2017) for intersemiotic relations in multimodal texts, and informed by Lim's (2021)

conceptualisation of co-contextualising and re-contextualising patterns. Three types of relationships were thus identified: Aesthetic Enhancement (AES), Humorous Engagement (HUM) and Semantic Support (SEM).

AES relationships were those where visual and verbal elements combined to enhance the recipient's positive face through beauty, elegance, or sensory appeal. In these cases, visual elements do not directly relate to gratitude concepts, but communicate messages of value and investment in the relationship, making the recipient feel special through premium visual treatment. For example, formal gratitude expressions like *Thank you*, presented in elegant calligraphy and framed by watercolour botanical garland, suggest the recipient deserves beautiful things. Similarly, multisensory aesthetic features such as embossing, glitter, or raised textures applied to written text enhance the material experience of receiving gratitude beyond purely linguistic expression.

HUM relationships emerged when visual and verbal modes collaborated to create wit and playfulness, requiring cognitive effort from recipients to decode intended meanings. This category encompasses two mechanisms: visual reinforcement of existing textual puns, where wordplay already present in written gratitude is amplified through corresponding imagery (e.g., *Thank you BERRY much* accompanied by berry illustrations, Card 71); and visual pun creation, where apparently straightforward gratitude text becomes humorous only through unexpected visual elements that create double meanings (e.g., *Thanks for the support* paired with images of bras, Card 4).

SEM relationships were characterised by logical connections between visual and verbal content, with images directly illustrating, reinforcing, or clarifying gratitude messages. This category was further subdivided into literal semantic support, where images provide concrete representations of abstract gratitude concepts (e.g., *you're a life saver*, accompanied by life preserver imagery, where the visual directly illustrates the rescue concept expressed verbally, Card 26), and figurative semantic support, where visuals help decode metaphorical or figurative verbal messages (e.g., *you make the world brighter*, supported by stars and sparkles that visually demonstrate the brightening effect described in the text, Card 91).

The above analytical framework enabled examination of how different semiotic modes interact, integrate and harmonise in thank-you cards to create what Stöckl and Bateman (2022) describe as “multimodal coherence” (p. 1).

3.3.4. Validation

We each independently tested the analytical framework we developed on 17 thank-you cards. Together, we then discussed ambiguous and hard-to-classify cases until we refined the labels and definitions of the various components of

our coding scheme, which we jointly applied to the entire corpus.

4. Findings

4.1. Findings of the linguistic analysis

4.1.1. The head acts

The 97 thank-you cards considered comprise 682 words, of which 572 encode the head acts and 110 the supportive moves. Therefore, each card contains a verbal text that is, on average, 7 words long. All the cards, except one (i.e., Card 25), have an explicit head act. Supportive moves, instead, are found in only 13 cards. Therefore, the head acts are much more prominent than the supportive moves, not only because they take up the bulk of the verbal text, but also because of their higher dispersion value.

The head act of the verbal speech act of thanking may consist of a single expression of gratitude plus its optional object(s) of the thanking illocution (i.e., the benefit(s)), or it may consist of a combination of two or more such expressions (i.e., a gratitude cluster; see Section 3.3.1). In this paper, to refer to an individual *token* of a gratitude expression plus its optional benefit, we use the term *head act*; instead, the part of the speech act where the illocution and its object(s) are expressed is called *head act component*. Thus, a head act component may comprise one or more head acts.

Short Text Cards (1-10 words)		Long Text Cards (11+ words)	
Word count	Number of cards	Word count	Number of cards
1	4	11	1
2	20	12	1
3	14	13	0
4	14	14	3
5	7	15	1
6	9	16	3
7	5	17	1
8	2	18	0
9	4	19	1
10	4	20	2
		30	1

Table 1
Word count distribution of the head act components.

In our corpus, a minority of the cards (i.e., 12) contain gratitude clusters (i.e., multiple head acts) in their head act components; for example, Card 8 contains two gratitude expressions (i.e., *A little note to say thank you. Thank*

you). Most, instead, contain single expressions of gratitude (e.g., *Thank you*, Card 21). The length of the head act components is quite variable, though most are very short (see Table 1). Indeed, most are between 2 and 6 words long; the shortest and the longest are, respectively, 1 and 30 words long. We considered word strings of letters separated by blank spaces, apostrophes, commas, full stops or exclamation marks. Thus, *I'm* (Card 90) counts as two words, but *every-fin* (a wordplay on *everything*, Card 30) counts as one, and so does *my* [picture of excrement] (Card 10), since what follows the possessive *my* is a picture, not a word.

The notion of gratitude is encoded in a few lexical realisations (see Table 2): *gratitude* (3 occurrences in 3 cards), *grateful* (3 in 3 cards), *thankful* (1), *thanks* (30 in 30 cards), *thank you* (80 occurrences in 60 cards), *appreciation* (once in 1 card), and *appreciated* (twice in 2 cards), while four more realisations are humorous variations on *thanks* (i.e., *fanks*, *frank*, *thankxolotl*) or *grateful* (i.e., *gratefull*), which appear to make sense when considered in relation to the visual component of the cards (see Section 4.3). Indeed, these differently spelt gratitude expressions pick up on, and mimic phonetically, (part of) the pictures in the cards, such as a dog's fangs with *Fanks* in Card 15, frankfurters with *Frank* in Card 27, a salamander with the last part of [*thankx*]olotl in Card 74.

Lexis	Occurrences
<i>Gratitude</i>	3
<i>Grateful</i>	3
<i>thanks</i>	30
<i>Thank you</i>	80
<i>Appreciation</i>	1
<i>Appreciated</i>	2
<i>Fanks</i>	1
<i>Frank</i>	1
<i>thankxolotl</i>	1
<i>gratefull</i>	1

Table 2
Lexical realisations of the notion of gratitude.

Lexical upgraders of gratitude expressions, coded as Intensifiers in our coding scheme, are found in 43 head acts, and only Cards 84 and 88 contain two intensifiers each.⁸ They are mostly realised with varied expressions of quantity (i.e., adverbials denoting the amount of gratitude being conveyed; e.g., *very much* in *Thank you very much*, Card 60), but also adverbs denoting the degree of intensity (e.g., *so* in *I'm so grateful*, Card 90). Infrequent

⁸ We did not consider other types of upgraders like repetitions of the gratitude expression, which we classified instead as a gratitude cluster, or the use of exclamation marks. For the use of typographic devices like all caps or bolding, see Section 4.1.3.

realisations include a noun (e.g., *an abundance of* in *An abundance of gratitude*, Card 1), three adjectival phrases denoting a great amount (e.g., *a very big* in *A little card to say a very big thank you*, Card 47), and a bound morpheme (i.e., *-olotl*, part of *Thankxolotl*, Card 74).

It is mainly in these internal modifiers of the head acts that one can notice the playful nature of many of these cards. Indeed, they provide an opportunity for plays on words such as *berry much* (Card 7), *so matcha* (Card 13), and *donut [...] enough* (Card 38). As in the case of funny gratitude expressions, these playful upgraders often rely on the accompanying pictures for a humorous effect. Thus, *berry much* is visually sustained by drawings of strawberries (Card 7); *so matcha* (Card 13) is accompanied by the drawing of a hand pouring green tea from a teapot into a cup; and *donut [...] enough* (Card 38) is paired with the drawing of an anthropomorphised, smiling doughnut soaking in a cup.

The complete list of the lexical intensifiers we identified is the following (see Table 3): *an abundance of*; *a brunch*; *a bunch* (4 times); *a huge*; *a latte*; *a melon* (twice); *a ton*; *a very big*; *beary much*; *berry much* (twice); *can't [...] enough*; *carrot [...] enough*; *doesn't [...] enough*; *donut [...] enough*; *ever so*; *for the record*;⁹ *from my head tomatoes* (with the letter strings *to*, *ma* and *toes* differently coloured); *many*; [...]olotl; *much* (once); *oh-so-much*; *so matcha*; *so* (3 times); *so much* (twice); *soup-er*; *soy much* (twice); *so very much*; *this much*; *very much* (3 times).

Intensifiers	Occurrences	Intensifiers	Occurrences
<i>An abundance of</i>	1	<i>Ever so</i>	1
<i>A brunch</i>	1	<i>For the record</i>	1
<i>A bunch</i>	4	<i>from my head tomatoes</i>	1
<i>A huge</i>	1	<i>many</i>	1
<i>A latte</i>	1	[...]olotl	1
<i>A melon</i>	2	<i>much</i>	1
<i>A ton</i>	1	<i>oh-so-much</i>	1
<i>A very big</i>	1	<i>so matcha</i>	1
<i>Beary much</i>	1	<i>so</i>	3
<i>Berry much</i>	2	<i>so much</i>	2
<i>Can't ... enough</i>	1	<i>soup-er</i>	1
<i>Carrot ... enough</i>	1	<i>soy much</i>	2
<i>Doesn't ... enough</i>	1	<i>so very much</i>	1
<i>Donut ... enough</i>	1	<i>this much</i>	1
		<i>very much</i>	1

Table 3
Lexical intensifiers of the notion of gratitude.

⁹ We considered *for the record* an upgrader because it is an idiom that is used to emphasise a point that the addressee should notice.

References to the benefits occur 34 times in 31 cards, and therefore exhibit similar frequency and dispersion values. Benefits include actions performed by the benefactor (17 occurrences; e.g., *and still you make time to show you care*, Card 96), qualities of, or roles played by, the benefactor (13 occurrences; e.g., *for simply being you*, Card 95), attitudes or experiences characteristic of the benefactor (1 occurrence; e.g., *for believing in me*, Card 12), or generic goods or services (3 occurrences; e.g., *for everything*, Card 35).

Excluding the formulation *Thank you*, and the presence of the possessive *your*, explicit references to the benefactor are rare, appearing 11 times in 8 cards. It may show up as the object of a preposition (e.g., *because of you*, Card 16), the subject of a clause (e.g., *all that you do*, Card 42), or a subject complement (e.g., *simply being you*, Card 95).

Only Card 28 contains an alerter, the humorous attention-getter *Hay* (a play on *Hey*).

The syntactic formulation of the head acts is varied but mostly conforms to the combination of the interjections *Thank you* or *Thanks* with a *to*-headed prepositional phrase denoting the benefactor or a *for*-headed prepositional phrase denoting the benefit, optionally enriched by an adverbial. More elaborate formulations are also attested, for instance when the benefit is encoded in a sentence or incomplete sentence. The various realisations are listed in Tables 4, 5 and 6.

Formulations	Frequency
<i>Thank you</i>	19
<i>Thank you</i> + <i>to</i> PP	1
<i>Thank you</i> + <i>for</i> PP	9
<i>Thank you</i> + other PP	1
<i>Thank you</i> + adverbial	12
<i>Thank you</i> + adverbial + <i>for</i> PP	1
<i>Thank you</i> + adverbial + PP	1
<i>Thank you</i> + adverbial + AP + <i>for</i> PP	2
<i>Thank you</i> + sentence	7
<i>Thank you</i> + incomplete sentence + NP	1
<i>Thank you</i> + sentence + <i>for</i> PP	2
<i>Thank you</i> + sentence + adverbial	1

Table 4
Formulations of head acts with the interjection *Thank you*.

Formulations	Frequency
<i>Thanks</i>	2
<i>Thanks + to PP</i>	0
<i>Thanks + for PP</i>	7
<i>Thanks + other PP</i>	0
<i>Thanks + adverbial</i>	12
<i>Thanks + adverbial + PP</i>	1
<i>Thanks + alerter</i>	1
<i>Thanks + incomplete sentence</i>	1

Table 5
Formulations of head acts with the interjection *Thanks*.

Other formulations	Frequency
Other interjection	1
Other interjection + adverbial	1
NP	1
Sentence	4
Sentence + NP	1
NP + to PP + incomplete sentence	1
Incomplete sentence	1
Incomplete sentence + sentence	1
NP + PP	2
AP + PP	1

Table 6
Other formulations of the head acts.

4.1.2. The supportive moves

Only 13 speech acts include single (10) or, more rarely, combinations (3) of supportive moves (see Table 7). These exemplify compliments to the thankee (7 PLI in 6 cards; e.g., *You're a life saver*, Card 26), sayings (4 PRO in 4 cards; e.g., *Life is crammed with things to take care of*, Card 96) and expressions of emotional impact (5 IMP in 5 cards; e.g., *Just thinking about your kindness brings a smile*, Card 94). Only Card 21 contains a humorous word-play (*You're otterly amazing*). These moves either enhance the perceived value of the benefit or positively evaluate the benefactor's behaviour or qualities. No instances were found of Acknowledgement of the imposition (ACK), Expression of indebtedness (DEB) or Promise of reciprocation (REP). Of the moves, one is encoded as a noun phrase (i.e., *Life saver*, Card 87), one as a coordinate clause (i.e., *and so are you*, Card 83), two as combinations of sentences (i.e., *Life is crammed with things to take care of. That means everything*, Card 96), and the rest as single sentences (e.g., *You make the world a better place*, Card 81).

Supportive moves	Frequency
ACK	0
DEB	0
IMP	5
PLI	7
PRO	4
REP	0

Table 7
Frequency of different types of supportive moves.

4.1.3. *Typographic characteristics of the verbal speech acts*

The use of typographic devices is summarised in Table 8. About half of the cards have verbal text characterised by a single typographic feature. Of the remaining half, 30.9% display two features, and 17.6% display three or more. More specifically, among the single typographic features (see Table 9), CAP is the most frequent one (almost 40%), followed by ITL, BOL, and STD, which display similar frequency values (i.e., around 20%). Among the double features (see Table 10), the most frequent is BOL+CAP (30%), while only three additional combinations (i.e., BOL+ITL, CAP+ITL, CAP+STD) account for just over 13% of the data each. Among the combinations of triple features (Table 11), only three occur more than once, namely BOL+CAP+COL, BOL+CAP+ITL and CAP+COL+STD. Finally, only one card shows a combination of four typographic features.

No. of typographic features	No. of cards
1	48 (49.5%)
2	31 (31.9%)
3	17 (17.6%)
4	1 (1.0%)
Total	97 (100%)

Table 8
Summary of typographic features of the verbal text.

Single typographic features	No. of cards
BOL	10 (20.8%)
CAP	18 (37.5%)
COL	0 (0.0%)
EBS	0 (0.0%)
GLI	0 (0.0%)
INC	0 (0.0%)
ITL	11 (22.9%)
STD	9 (18.8%)
Total	48 (100%)

Table 9
Frequency of single typographic features in the verbal text.

Double typographic features	No. of cards
BOL + CAP	9 (29.0%)
BOL + COL	1 (3.2%)
BOL + ITL	4 (12.9%)
BOL + STD	1 (3.2%)
CAP + COL	2 (6.5%)
CAP + ITL	5 (16.1%)
CAP + STD	4 (12.9%)
COL + ITL	2 (6.5%)
COL + STD	1 (3.2%)
ITL + STD	2 (6.5%)
Total	31 (100%)

Table 10
Summary of double typographic features of the verbal text.

Triple typographic features	No. of cards
BOL + CAP + COL	4 (23.5%)
BOL + CAP + ITL	4 (23.5%)
BOL + CAP + STD	1 (5.9%)
BOL + COL + ITL	1 (5.9%)
BOL + ITL + STD	1 (5.9%)
CAP + COL + ITL	1 (5.9%)
CAP + COL + STD	3 (17.6%)
CAP + ITL + STD	1 (5.9%)
COL + ITL + STD	1 (5.9%)
Total	17 (100%)

Table 11
Summary of triple typographic features of the verbal text.

4.2. Findings of the visual analysis

The analysis of visual and verbal content distribution showed that front-only distribution (FRONT) dominates with 80 cards, where images and written text appear exclusively on the front cover, leaving the interior of the card blank. Extended distribution (EXTENDED) accounts for 16 cards, where content spans both the front and the interior of the card. Only one card (Card 93) employs progressive distribution (PROGRESSIVE), where visual and textual elements continue a narrative from the front to the interior.

The visual analysis of the 97 thank-you cards revealed a variety of subjects depicted both as single visual elements and in combinations (see Table 12). Cards with single visual elements display food imagery as the most prominent category (21 occurrences), followed by floral/botanical elements (20 occurrences) and animals (18 occurrences). Abstract/geometric patterns appear in 8 cards, objects in 11, typography-only designs in 2, scenes in 2, and people in Card 60 only. Additionally, 14 cards display more elaborate compositions combining multiple visual elements. Compositions of objects and floral/botanical elements (vase and flowers) appear in 5 cards;

combinations of landscapes and background scenes with flowers or animals or objects are found in 6 cards; 2 cards combine abstract elements with animals or flowers, and Card 40 displays a complex combination of abstract elements, flowers, animals, and objects.

Subjects depicted	Frequency
FOO	21
FLO	27
ANI	24
OBJ	20
ABS	10
TYP	2
SCE	8
PEO	1

Table 12
Frequency of subjects depicted in the cards.

The food category encompasses a diverse range of items, including anthropomorphised pastries (Card 1), ice cream (Card 2), fruits like grapes (Card 5) and strawberries (Card 7), beverages such as coffee and tea (Cards 13, 55), and other prepared foods. Animal imagery spans domestic pets including cats (Cards 19, 35, 86) and dogs (Cards 15, 29), farm animals including pigs (Card 31) and horses (Card 28), wildlife including bears (Cards 24, 79), sea otters (Card 21), and other marine life (Cards 11, 30), and bees (Cards 10, 11). Botanical elements range from simple decorative borders (Cards 37, 46) to elaborate watercolour arrangements (Cards 39, 83) and to wildflower compositions (Cards 45, 48).

The analysis of content revealed that 76 cards display conceptual representations (CON) showing static objects, symbols, or decorative elements. Narrative actions (NAR) appear in 21 cards, depicting dynamic scenes where characters (mainly anthropomorphized objects, animals or food items) perform actions or engage in human-like behaviours.

Stylistic rendering analysis showed a strong preference for stylized/cartoon/simplified imagery (STY) appearing in 83 cards, while realistic/photographic representation (REA) occurs in 14 cards. This stylistic choice suggests a tendency toward playful, accessible visual communication rather than formal photographic presentation.

Front-page visual analysis revealed that central positioning with high visual importance (CEN) appears in 89 cards, making it the preferred layout choice. Marginal or decorative positioning (MAR) occurs in 8 cards, typically for subtle accent elements. Frame characteristics showed strong borders or frames (FRA) in 94 cards, while open, boundary-free compositions (OPE) appear in only 3 cards.

Interior pages analysis showed a marked preference for marginal

positioning (MAR) of visual elements with 16 occurrences versus 1 occurrence for central positioning (CEN) in Card 94. This reversal of compositional priorities indicates that interior visual elements serve supporting functions, mainly displaying decorative frames that enclose the written text and echo the pattern, colour palette, or visual elements presented on the front cover.

4.3. Findings of the intersemiotic analysis

The intersemiotic analysis revealed three primary relationship types between visual and verbal elements: HUM, AES and SEM (see Table 13).

Intersemiotic relationship	Frequency
HUM	44
AES	32
SEM	11
HUM+SEM	2
AES+SEM	8
Total	97

Table 13
Frequency of intersemiotic relationships.

HUM appears in 44 cards, making it the most frequent relationship type. This category encompasses visual reinforcement of textual puns, such as Card 7's *thank you berry much* accompanied by anthropomorphised strawberries with happy facial expressions, or Card 13's *Thanks so MATCHA* paired with a tea preparation scene showing matcha being poured into a cup. Intersemiotic wordplay occurs where written text becomes humorous through unexpected imagery, as in Card 4's *Thanks for all the support* depicted with various colourful bras.

AES occurs in 32 cards, where visual elements enhance the reader's positive face through beauty and elegance without additional semantic layers. These cards feature refined presentations such as Card 37's simple *Thank you* framed by watercolour floral borders in purple, blue, and orange tones, or Card 41's *Thank you for your kindness* accompanied by a wildflower bouquet in a glass jar.

SEM appears in 11 cards as a standalone function. However, semantic interplay combines with other relationship types in 10 additional cases: 2 cards display both SEM and HUM, while 8 cards combine SEM with AES.

Standalone SEM includes literal semantic support, for instance, Card 20's *Thanks a million* illustrated with a red wheelbarrow filled with green paper money, directly representing the large quantity concept. Figurative SEM appears in Card 16: *The wheels go round because of you. Thank you!*

combines with the image of a yellow school bus, connecting the figurative “wheels” of organization to literal vehicle wheels.

Combined semantic functions demonstrate how visual elements can serve multiple purposes simultaneously. Card 87’s *Life saver* accompanies an orange and white life ring on blue ocean waves. This exemplifies the combination of SEM+HUM where the life ring imagery clarifies the rescue metaphor (SEM) and creates wordplay humour through the visual reference to life ring (HUM). Card 52’s *Thanks a bunch* pairs with a watercolour bouquet exemplifying SEM+AES interplay, where the floral imagery both clarifies the double meaning of *bunch* (SEM), while providing aesthetic enhancement through botanical illustration.

The distribution of these relationship types correlates with visual element categories: food imagery most frequently serves humorous functions, floral elements predominantly provide aesthetic enhancement, while animal imagery demonstrates versatility across all three relationship types, reflecting their adaptability in multimodal meaning construction.

5. Discussion and conclusion

5.1. Discussion of the findings

The linguistic analysis shows that the verbal component of the cards is characterised by the very frequent presence of explicit and prototypical head acts of gratitude. Indeed, most of the gratitude-related expressions are based on the morpheme *thank*. This suggests that the spatial constraints and face-enhancing function of thank-you cards may favour verbal expressions that are immediately clear as to the sentiment being conveyed. Only a few creative, playful adaptations (e.g., *fanks*, *thankxolotl*) are attested, which can be easily understood when anchored to the visual mode. This shows, on one hand, that the verbal and visual modes of expression are complementary and co-dependent, and on the other, that humour operates as an auxiliary resource for affective engagement.

Lexical upgraders modify over 40% of the head acts. They include conventional intensifiers (*very much*, *so*), and inventive ones derived from plays on words (*berry much*, *so matcha*), reinforced by the visual component. This suggests a synergistic relationship between lexis and image, where the former hints at humour or affection, which can only be fully processed when considered together with the latter. The frequency of playful engagement as the primary visual-verbal relationship reveals humour’s central role in transforming gratitude into relational bonding. The cards achieve this through intersemiotic wordplay, which requires cognitive collaboration between sender and recipient to decode visual-verbal puns. This shared interpretive

work points to an investment in the relationship beyond mere acknowledgement of benefits received.

Supportive moves are not only infrequent, but also restricted to compliments, general reflections, and statements of emotional impact. The cards thus appear to be verbally focused on positive affect and celebration rather than the negotiation of social obligations, serving more to boost than to redress social relations. This verbal restraint, combined with the predominance of AES and HUM over SEM, suggests that the cards serve as a means for managing social relationships.

Typographic features are a pervasive feature of the cards, used with subtle variations, both as single formatting choices and in combinations. The most common is CAP, followed by BOL and ITL. They add visual salience to the textual component, and may be considered a form of visual emphasis. The frequent combinations of multiple typographic resources not only reinforce the multimodal nature of the card in general, but also specifically draw the reader's attention to the *verbal* component of the cards, which might otherwise go unnoticed. This typographic strategy works in concert with compositional choices that position visual elements centrally for immediate impact, while ensuring the verbal gratitude expressions remain prominent. This reflects the cards' dual function of providing visual pleasure while maintaining clear communicative intent.

While the mutual reinforcement of words and images is frequently attested, the actual replacement of words with pictures occurs only twice (Cards 22 and 51) as a euphemistic device to avoid using four-letter words. This gives rise to multimodal sentences (cf. D'Ulizia *et al.* 2007) in which the subject and predicator are expressed verbally, while the direct object or adverbial are rendered visually.

In general, the visual components in multimodal thank-you cards serve informational-functional roles that extend far beyond mere decoration or illustration. They operate as active meaning-making resources that disambiguate verbal wordplay, amplify abstract gratitude concepts through concrete imagery, and intensify emotional engagement through aesthetic appeal and anthropomorphised characters.

Overall, the description of the verbal component suggests that the cards are lexically and typographically optimised for immediacy and clarity, while intersemiotic relationships and visual aesthetics serve as supplementary resources for emotional engagement, producing humour and enhancing visual appeal.

The analytical framework recognises that thank-you cards employ multiple strategies for coordinating visual and verbal elements, ranging from simple reinforcement to complex interplay, that create sophisticated pragmatic effects. Our analysis reveals that multimodal texts cannot be fully understood through additive approaches that merely combine verbal and

visual meaning. Rather, intersemiotic relationships create emergent pragmatic effects that exceed the sum of individual modal contributions; as a result, the cards' effectiveness depends on readers' ability to integrate multiple semiotic resources into coherent communicative experiences.

5.2. Concluding remarks

The integrated linguistic, visual and intersemiotic analysis has provided a holistic representation of a small corpus of thank-you cards, showing that typographic prominence, concise explicit gratitude, and visually anchored humour are key communicative strategies for engaging audiences.

The brevity of the texts, the conventionality of their gratitude expressions, and the limited number of their supportive moves suggest that the primary goal of such cards is to convey the message at a glance, rather than elaborate on it. Also, the occasional recourse to humour indicates that an additional function of such a message is to provide entertainment (i.e., aesthetic appreciation), possibly as a form of gift-giving or reciprocation meant to please the addressee further.

Our analysis also reveals that politeness strategies operate through intermodal relations. While verbal elements maintain clarity and immediacy through explicit expressions of gratitude, visual elements serve affective and relational goals through aesthetic enhancement and humour creation. The affordances of individual modes and their co-deployment allow the cards to ensure comprehensible illocutionary force and to maximise emotional engagement.

This study presents a few limitations. First, the corpus we analysed comprises exclusively English-language cards, limiting our understanding of how the multimodal properties identified might manifest across different linguistic and cultural contexts.

Second, the data collection process introduced potential biases that may have influenced the representativeness of the corpus. The cards collected through personal networks reflect the aesthetic preferences and social demographics of the collectors, while the *Etsy* selection process, guided by platform algorithms and search parameters, probably favoured commercially popular designs over niche or experimental approaches. Additionally, the time gap between the personal and online collections of the cards precluded the examination of a unified time period, potentially introducing variations related to changing design trends or seasonal factors.

Third, the corpus size of 97 cards, while sufficient for identifying prominent patterns, does not allow us to make definitive claims about frequency distributions or rare phenomena within the genre. A larger dataset might reveal different proportions of intersemiotic relationship types or uncover additional multimodal strategies not represented in this sample.

Future research could examine thank-you cards from additional sources, in other languages, and across a broader price range to determine their degree of genre-internal variation. Additionally, in line with Abdullah and Nor (2022), investigations could be conducted into the digital manifestations of gratitude, such as e-cards and memes, to examine whether and how they replicate or adapt the multimodal strategies observed in print cards, including verbal-visual wordplay and expressive typography. Furthermore, reception studies could be conducted to investigate whether readers notice the content, humour, typographic emphasis, and visual characteristics of thank-you cards, and how they cognitively and emotionally respond to them. Finally, researchers could analyse cards associated with other speech acts, such as apologies, congratulations, or condolences, to identify both their shared and distinctive features.

Bionotes: Elisa Bertoldi is a postdoctoral researcher at the University of Padua. Sara Gesuato researches and teaches English linguistics at the University of Padua.

Authors' addresses: elisa.bertoldi.1@unipd.it, sara.gesuato@unipd.it

Acknowledgements: We thank two anonymous reviewers for helpful comments on our paper.

References

- Abdullah S. and Nor J.M. 2022, *Exploring Non-linguistic Features of Malay Hari Raya E-cards*, in “International Journal of Academic Research in Business and Social Sciences” 12 [1], pp. 1076-1083.
- Ahar V. and Eslami-Rasekh A. 2011, *The Effect of Social Status and Size of Imposition on the Gratitude Strategies of Persian and English Speakers*, in “Journal of Language Teaching & Research” 2 [1], pp. 120-128.
- Al-Ali M.N. 2006, *Conveying Academic and Social Identity in Graduate Dissertation Acknowledgements*, in Neumann C.-P., Pérez-Llantada Auría C. and Alastruér R.P. (ed.) *Proceedings of the Fifth International AELFE Conference*, Prensas Universitarias de Zaragoza, University of Zaragoza, pp. 35-42.
- Al-Subhi A.S. 2024, *Interactional Multimodal Metadiscourse in Public Health Posters During The COVID-19 Pandemic*, in “Pragmatics & Society” 16 [2], pp. 255-281.
- Atminytė E. 2022, *Soviet Lithuanian New Year Greeting Card: Semiotics of Recto*, in “Semiotika” 17, pp. 162-185.
- Ben-Ari E. 1987, *On Acknowledgements in Ethnographies*, in “Journal of Anthropological Research” 43 [1], pp. 63-84.
- Blum-Kulka S., House J. and Kasper G. 1989, *Cross-cultural Pragmatics: Requests and Apologies*, Ablex Publishing Corporation, Norwood (NJ).
- Bouko C. 2020, *Emotions through Texts and Images: A Multimodal Analysis of Reactions to the Brexit Vote on Flickr*, in “Pragmatics” 30 [2], pp. 222-246.
- Brown P. and Levinson S.C. 1978, *Universals in Language Use: Politeness Phenomena*, in Goody E.N. (ed.), *Questions and Politeness: Strategies in Social Interaction*, Cambridge University Press, Cambridge, pp. 56-289.
- Cheng S.W. 2005, *An Exploratory Cross-sectional Study of Interlanguage Pragmatic Development of Expressions of Gratitude by Chinese Learners of English*, PhD dissertation, University of Iowa.
- Cohn N., Murthy B. and Foulsham T. 2017, *A Picture is Worth More Words Over Time: Multimodality and Narrative Structure Across Eight Decades of American Superhero Comics*, in “Multimodal Communication” 6 [1], pp. 1-18.
- Coulmas F. 1981, “Poison to your soul”. *Thanks and Apologies Contrastively Viewed*, in Coulmas F. (ed.), *Conversational Routine. Explorations in Standardized Communication Situations and Prepatterned Speech*, Mouton, The Hague/Paris/New York, pp. 69-91.
- Cronin B. 1995, *The Scholar’s Courtesy. The Role of Acknowledgement in the Primary Communication Process*, Taylor Graham, London/Los Angeles.
- Del Saz-Rubio M.M. 2018, *Female Identities in TV Toiletries Ads: A Pragmatic and Multimodal Analysis of Implied Meanings*, in “Journal of Pragmatics” 136, pp. 54-78.
- Díaz Pérez F.J. 2005, *The Speech Act of Thanking in English: Differences Between Native and Non-native Speakers’ Behaviour*, in “Revista de Filología Inglesa” 26, pp. 91-102.
- D’Ullizia A., Ferri F. and Grifoni P. 2007, *A Hybrid Grammar-based Approach to Multimodal Languages Specification*, in Meersman R. et al. (ed.), *On the Move to Meaningful Internet Systems 2007: OTM 2007 Workshops. OTM 2007. Lecture Notes in Computer Science*, Springer-Verlag, Berlin/Heidelberg, pp. 367-376.
- Eisenstein M. and Bodman J.W. 1986, “I very appreciate”: *Expressions of Gratitude by Native and Non-native Speakers of American English*, in “Applied Linguistics” 7

- [2], pp. 167-185.
- Eisenstein M. and Bodman J.W. 1993, *Expressing Gratitude in American English*, in Kasper G. and Blum-Kulka S. (ed.), *Interlanguage Pragmatics*, Oxford University Press, New York/Oxford, pp. 64-81.
- Etsy. 2025, *Etsy – Global Digital Marketplace*, <https://www.etsy.com/> (last accessed January 2025).
- Forceville C. and Kjeldsen J.E. 2018, *The Affordances and Constraints of Situation and Genre: Visual and Multimodal Rhetoric in Unusual Traffic Signs*, in “International Review of Pragmatics” 10 [2], pp. 158-178.
- Francesconi S. 2011, *Multimodally Expressed Humor Shaping Scottishness in Tourist Postcards*, in “Journal of Tourism and Cultural Change” 9 [1], pp. 1-17.
- García O. and Kleifgen J.A. 2016, *Translanguaging and Literacies*, in “Reading Research Journal” 0 [0], pp. 1-19.
- Giannoni D.S. 1998, *The Genre of Journal Acknowledgements: Findings of a Crossdisciplinary Investigation*, in “Linguistica e Filologia” 6, pp. 61-84.
- Giannoni D.S. 2002, *Worlds of Gratitude: A Contrastive Study of Acknowledgement Texts in English and Italian Research Articles*, in “Applied Linguistics” 23 [1], pp. 1-31.
- Giannoni D.S. 2006, *Book Acknowledgements Across Disciplines and Texts*, in Hyland K. and Bondi M. (ed.), *Academic Discourse Across Disciplines*, Peter Lang, Berlin, pp. 151-175.
- Gillen, J. 2013, *Stoic and Sentimental: The Emotional Work of the Edwardian Greetings Postcard*, in “Journal of Social History” 47 [2], pp. 313-341.
- Goffman E. 1981, *Forms of Talk*, University of Pennsylvania Press, Philadelphia.
- Guo F.B. and Feng X. 2017, *A Multimodal Discourse Analysis of Advertisements-based on Visual Grammar*, in “Journal of Arts and Humanities” 6 [3], pp. 59-69.
- Held G. 1996, *Two Polite Speech Acts in Contrastive View: Aspects of the Realization of Requesting and Thanking in French and Italian*, in Hellinger M. and Ammon U. (ed.), *Contrastive Sociolinguistics*, Mouton de Gruyter, Berlin/New York, pp. 363-384.
- Hidalgo-Downing L. and Kraljević Mujic B. 2015, *Multimodal Metaphor and Fictional Storytelling in Printed and Internet Ads: Recontextualizing Social Practices and Globalization*, in “Revista Brasileira de Linguística Aplicada” 15, pp. 377-402.
- Hiippala T., 2007, *Linguistic and Multimodal Insights into Tourist Brochures*, in “Finnish Journal of Tourism Research”, 10 [2], pp. 62-68.
- Hyland K. 2003, *Dissertation Acknowledgements: The Anatomy of a Cinderella Genre*, in “Written Communication” 20 [3], pp. 242-268.
- Hyland K. 2004, *Graduates’ Gratitude: The Generic Structure of Dissertation Acknowledgements*, in “English for Specific Purposes” 23, pp. 303-324.
- Hyland K. and Tse P. 2004, *‘I would like to thank my supervisor’. Acknowledgements in Graduate Dissertations*, in “International Journal of Applied Linguistics” 14 [2], pp. 259-275.
- Kress G. and van Leeuwen T. 2021, *Reading Images. The Grammar of Visual Design*, 3rd ed., Routledge, London/New York.
- Lambeens T. and Pint K. 2015, *The Interaction of Image and Text in Modern Comics*, in Lardinois A., Levie S., Hoeken H. and Lüthy Ch. (ed.), *Texts, Transmissions, Receptions: Modern Approaches to Narratives*, Brill, Leiden, pp. 240-257.
- Lim F.V. 2021, *Investigating Intersemiosis: A Systemic Functional Multimodal Discourse Analysis of the Relationship Between Language and Gesture in Classroom Discourse*, in “Visual Communication” 20 [1], pp. 34-58.
- Liu Y. 2023, *Analysing Tension Between Language and Images: A Socialsemiotic View*, in

- “Social Semiotics” 33 [1], pp. 132-150.
- Liu Y. and O’Halloran K. L. 2009, *Intersemiotic Texture: Analyzing Cohesive Devices Between Language and Images*, in “Social Semiotics” 19 [4], pp. 367-388.
- McWhorter J. 2012, *A Surprising New Language – Texting* <https://www.youtube.com/watch?v=yoF2vdLxsVQ> (8.8.2025).
- Moya-Guijarro A.J. 2016, *A Multimodal Analysis of Picture Books for Children: A Systemic Functional Approach*, Equinox, Sheffield.
- Nikolajeva M. and Scott C. 2006, *How Picturebooks Work*, Routledge, London and New York.
- Oakey D., Jones C. and O’Halloran K.L. 2022. *Phraseology and Imagery in UK Public Health Agency COVID-19 Tweets*, in O’Halloran K.L., Jones C. and Oakey D. (ed.), *Discourses, Modes, Media and Meaning in an Era of Pandemic*, Routledge, London, pp. 89-114.
- Okamoto S. and Robinson W.P. 1997, *Determinants of Gratitude Expressions in England*, in “Journal of Language and Social Psychology” 16 [4], pp. 411-433.
- Padlet. 2025, *Padlet – Online Collaboration Platform*, <https://it.padlet.com> (last accessed January 2025).
- Schauer G.A., Adolphs S. 2006, *Expressions of Gratitude in Corpus and DCT Data: Vocabulary, Formulaic Sequences, and Pedagogy*, in “System” 34, pp. 119-134.
- Searle J.R. 1976, *A Classification of Illocutionary Acts*, in “Language in Society” 5 [1], pp. 1-23.
- Sipe L.R. 1998, *How Picture Books Work: A Semiotically Framed Theory of Text-Picture Relationship*, in “Children’s Literature in Education” 29 [2], pp. 97-108.
- Skwarzyński J. 2019, *Reading Images. Comics and Its Multimodality in Cultural Communication, Interpretation and Translation*, in “New Horizons in English Studies” 4, pp. 102-117.
- Stöckl H. 2024, *Detecting Generic Patterns in Multimodal Argumentation: A Corpus-based Study of Environmental Protection Print-Ads*, in “Journal of Argumentation in Context” 13 [2], pp. 260-291.
- Stöckl H. and Bateman, J.A. 2022, *Editorial: Multimodal Coherence Across Media and Genres*, in “Front. Commun” 7 [1104128], pp. 1-2.
- Tseronis A. 2015, *Multimodal Argumentation in News Magazine Covers: A Case Study of Front Covers Putting Greece on the Spot of the European Economic Crisis*, in “Discourse, Context & Media” 7, pp. 18-27.
- Unsworth L. and Cléirigh C. 2017, *Multimodality and Reading: The Construction of Meaning Through Image-Text Interaction*, in Jewitt C. (ed.), *The Routledge Handbook of Multimodal Analysis*, 2nd edn., Routledge, London, pp. 176-188.
- Vásquez C. and Aslan E. 2021, “Cats be outside, how about meow”: *Multimodal Humor and Creativity in an Internet Meme*, in “Journal of Pragmatics” 171, pp. 101-117.
- Yu H. 2019, *One Page, Two Stories: Intersemiotic Dissonance in a Comic Adaptation of Journey to the West*, in “Social Semiotics” 31 [2], pp.1-21.

MULTIMODAL PRAGMATICS IN INTERCULTURAL COMMUNICATION Egyptian Multimodal Cues in English-mediated Interactions

LUCIA LA CAUSA
UNIVERSITY OF CATANIA

Abstract – This study explores the dynamics of multimodal communication in intercultural contexts, focusing on Egyptian users of English. Multimodal cues, such as gesture, eye contact, intonation, volume, and interpersonal distance, are culturally shaped, making cross-cultural communication prone to misunderstandings. Drawing on a multimodal pragmatic framework and a qualitative methodology based on Norris’s (2004a) Multimodal (Inter)action Analysis (MIA), the study examines a manually compiled corpus of YouTube videos featuring the Egyptian actor Bassem Youssef speaking English in anglophone environments. Specifically, through this case study, key para-linguistic and extra-linguistic resources are analysed to explore how Egyptian communicative style interacts with English-mediated discourse. The analysis investigates whether pragmatic transfer from Egyptian norms enhances communicative effectiveness or produces unintended impoliteness and challenges in cooperation. The case study’s findings offer empirical insight into culturally embedded Egyptian nonverbal practices in English interactions, highlighting the importance of cultural awareness in interpreting multimodal signals.

Keywords: communicative style; intercultural pragmatics; multimodal pragmatics; nonverbal communication; Egyptian English speakers.

1. Introduction

Communication is a dynamic process involving the interplay of verbal and nonverbal modes that jointly co-construct meaning in interaction (Balboni 2015; Kress, van Leeuwen 2001). As Norris (2004a) observes, “[a]ll interactions are multimodal” (p. 1), relying not only on linguistic resources but also on a wide range of paralinguistic (e.g., turn-taking, intonation, and rhythm) and extra-linguistic (e.g., gesture, gaze, facial expressions, body and head movement) semiotic resources. While language is often assumed to carry the most informative content, research demonstrates that para-linguistic and extra-linguistic modalities play equally crucial roles in meaning-making (see Indarti 2024; Kecskés 2014; Martinec 2000; Norris 2004a; O’Halloran *et al.* 2014; Payratò 2017).

However, decoding nonverbal meanings is not always immediate, even among interlocutors who share a linguistic and cultural background, and it becomes considerably more complex in intercultural interactions. As Kecskés (2014) argues, effective communication in such settings requires more than linguistic competence, including knowledge of vocabulary and grammar; it also demands intercultural pragmatic competencies, or the ability to interpret and produce culturally appropriate non-linguistic signals (see also Balboni 2006, 2015). He also observes that different speech communities often employ distinctive multimodal features to convey meanings, since these are deeply embedded in the speaker's native sociocultural background and pragmatic norms (see also Ting-Toomey 1999). When interlocutors do not share the same symbolic and cultural system, intercultural misinterpretations are likely to occur due to differences in values and socio-cultural norms, leading to miscommunication, pragmatic failure, or the perception of impoliteness (Abu-Arqoub and Alserhan 2019; Handriani *et al.* 2025; Mühleisen 2011). Even when English is used as a shared language and speakers use grammatically correct verbal elements, culturally rooted multimodal norms from a speaker's first language (L1) may influence interaction, potentially shaping communicative effectiveness (Kecskés 2014).

Through the illustrative case of the Egyptian actor and satirist Bassem Youssef, the present study investigates the multimodal practices of Egyptian users of English in English-mediated contexts. The main aim is to verify whether Egyptians' communicative style, widely characterised as expressive, animated, and high-contact (Nydell 2018), is transferred and may be misinterpreted in intercultural settings, without interlocutors' intercultural awareness or cultural sensitivity and adaptability. Adopting a multimodal pragmatic approach combined with Norris's (2004a, 2004b, 2011, 2020, 2021) Multimodal (Inter)action Analysis (MIA), this study examines a corpus of YouTube videos featuring Youssef communicating in English with English native speakers, in anglophone contexts (e.g., American or British interviews and podcasts). Focus is placed on multimodal cues – including some prosodic features, visual cues, interpersonal distance, eye contact, and touch – to investigate how Egyptians negotiate meaning in such English-mediated interactions. The resulting corpus is then qualitatively analysed to assess whether communication is pragmatically effective or whether the transfer of Egyptian norms results in misunderstanding or unintended social effects.

By documenting the Egyptian actor's nonverbal practices and offering empirical insights into how he negotiates meaning in English-mediated interactions, this investigation also wishes to contribute to the understanding of multimodal communication and intercultural pragmatics, reinforcing the theoretical integration of these two approaches.

2. Multimodal pragmatics in intercultural communication: State of the art

Multimodal pragmatics examines how language interacts with other non-linguistic communicative modes, including visual, gestural, auditory, and spatial modes, to serve pragmatic functions, such as co-constructing and reinforcing meaning, conveying intentions, managing politeness, and regulating turn-taking in interaction (O'Halloran *et al.* 2014; Payratò 2017). As Schroeder (2017) reports, the foundational relevance of multimodal communication was already recognised as early as classical antiquity, but only began to be systematically explored in the mid-twentieth century. Today, the literature is extensive, intersecting with multiple disciplines – such as social psychology, anthropology, semiotics, cognitive communication studies, sociolinguistics, and linguistic multimodal pragmatics – and encompassing key studies such as those by Bavelas (1990), Kendon (1967, 2004a), McNeill (1992, 2005), Kress and van Leeuwen (2001), O'Halloran *et al.* (2014), Payratò (2017), and Holler (2025), among others. However, early research primarily examined multimodal behaviours within a single culture and native-language contexts, leaving a notable gap in the intersection of multimodal and intercultural communication.

Still today, the exploration of how multimodal elements operate across linguistic and cultural boundaries remains in its early stages. Few investigations have explicitly examined how verbal and nonverbal modes interact in cross-cultural communication scenarios. Seminal contributions, such as Gumperz's ([1982] 2009) on prosody and gesture, and Tannen's ([1984] 2005) on conversational style, showed how intonation, rhythm, pitch, overlap, and gesture contribute to meaning and can generate misunderstanding among speakers from different cultural backgrounds. Hall's (1966, 1974, 1976) works on proxemics further highlighted that nonverbal behaviour is culturally embedded, rather than universal, a point widely supported by later research on intercultural communication. More recent studies, notably Kecskés (2014, 2022), link intercultural pragmatics with multimodality, underlining how pragmatic transfer from the L1 shapes both verbal and nonverbal behaviours in English as a Lingua Franca (ELF), sometimes resulting in miscommunication. Parallel developments in multimodal pragmatics and interaction analysis (Norris 2004a, 2011, 2020) offer systematic frameworks for examining how gaze, gesture, prosody, and spatial orientation coordinate with spoken language to produce meaning in intercultural interactions.

Yet, despite such valuable contributions, studies explicitly bridging multimodal and intercultural pragmatics remain scarce. Moreover, most research focuses on native-language interactions or language-learning contexts (see Beltrán-Planques and Querol-Julián 2018; Gumperz [1982]

2009; Ishihara 2018; O'Halloran *et al.* 2014; Tannen [1984] 2005, among others). In contrast, there remains a relative lack of empirical investigation into how speakers employ prosodic and nonverbal cues when communicating in a non-native language (e.g., English as a Lingua Franca), in non-native, transnational settings (see Kecskés 2014, 2022; Lopriore and Grazzi 2016). This gap is striking given English's global role as a medium of intercultural communication, especially in digital spaces. Being at the intersection of multimodal pragmatics and intercultural communication, and integrating insight from sociolinguistics, anthropology, cognitive psychology, social semiotics, and interactional discourse studies, the present study addresses this gap by empirically analysing how Egyptians use multimodal resources to negotiate meaning in English-mediated interactions, focusing on contexts such as online interviews and podcasts within anglophone cultural and linguistic settings.

3. Method

Egyptian communication is notably described as colourful, marked by emotionally nuanced tones, frequent hand gestures, animated facial expressions, strong eye contact, close interpersonal distance, and physical touch (Nydell 2018). These behaviours are culturally rooted and reflect broader social values such as warmth, hospitality, emotional engagement, and solidarity (Feghali 1997; Nydell 2018). However, while culturally normative within Egypt, such behaviours may be misread and perceived as overly assertive, inappropriate, or even disrespectful when displayed in intercultural contexts, sometimes resulting “in misunderstanding, leaving both parties feeling bewildered or deceived” (Nydell 2018, p. 93).

Because nonverbal communication is culturally specific (see Section 1), the expectation is that Egyptians continue to exhibit these patterns in English interactions. Even with high fluency, their communicative style may blend global, English-mediated verbal forms with local nonverbal practices. The resulting English expressive communication style may diverge from the pragmatic expectations of English-speaking interlocutors, especially those from cultures prioritising verbal restraint and minimal nonverbal display. To investigate these dynamics, this study explores Egyptian users of English in cross-cultural settings beyond Egypt, an underexplored area. While previous research has documented Egyptian nonverbal practices in Arabic-speaking environments (see Adly 2021; Eladawy 2021; Feghali 1997; Hafez 1991; Nydell 2018; Raddawi 2015; Zaharna 1995, among others), few or no studies address how these practices persist in English-mediated, anglophone settings. The research adopts a multimodal pragmatic approach combined with video-based qualitative analysis, building on Norris's (2004a, 2004b, 2011, 2020,

2021) MIA. The corpus was manually compiled from publicly available videos on YouTube, which integrate audio, visual, and textual data and enable detailed multimodal transcription and analysis of how verbal and nonverbal modes interact (Norris 2019; Haryanti *et al.* 2023). The selection process yielded twenty-one videos. Criteria primarily included the presence of fluent Egyptian English speakers interacting with native English interlocutors; semi-formal settings such as interviews, debates, or podcasts; face-to-face interaction (even though mediated by microphones); and clear visibility of the speaker's upper body to allow observation of gestures, body movement, facial expressions, and proxemics. The resulting corpus prominently features interviews with public figures, mainly actors and writers. Out of the twenty-one videos, seven were chosen for this analysis, consisting exclusively of interviews with Bassem Youssef, whose frequent use of paralinguistic and extralinguistic cues typical of Egyptian speakers, high fluency in English, and frequent participation in English-language (mainly American) shows and podcasts provide a rich and illustrative case-study for examining the persistence of Egyptian pragmatic norms in intercultural settings.

Data were analysed through repeated observation, including slow-motion viewing, and manual transcription in Word. Multimodal transcription conventions (Norris 2004a, 2011, 2019, 2020) were employed, integrating verbal utterances with paralinguistic and nonverbal signals. In addition, snapshots were extracted to document gestures, body movement, and interpersonal distance. Specifically, the analysis focused on six of Andersen's (2010) eight nonverbal codes: chronemics (turn-taking, rhythm), vocalics (prosody, intonation, volume), kinesics (body motion, gestures, facial expressions), proxemics (interpersonal distance), haptics (touch), and oculosics (eye contact). Physical appearance (gender, height, weight, skin colour, body shape, and artifacts) and olfactics (smell) were excluded, as they were not deemed relevant or applicable to the aims of this research. The primary aim was to identify patterns of pragmatic transfer and trace how Egyptian norms influence English communication. Particular attention was given to instances of miscommunication, difficulties with cooperation, perceptions of impoliteness, or unintended offence, which can make intercultural communication ineffective or pragmatically unsuccessful.

4. Results

4.1. Para-linguistic features

4.1.1. Turn-taking, rhythm, intonation, and volume

Chronemics and vocalics – including turn-taking, speech tempo, intonation, and volume – are culture-specific conversational resources (Balboni 2006; Stivers *et al.* 2009). While English-speaking cultures generally privilege orderly sequencing, Egyptian interactional norms value immediacy, involvement, and expressive intensity (Nydell 2018). Such contrasts often generate tension and misunderstanding in intercultural exchanges (Balboni 2006).

Although similarities between Arabic and English have been detected (Ali Aldhahri 2024), in many English-speaking settings, turn-taking typically follows a one-at-a-time model, where a speaker holds the floor until yielding it (Sacks *et al.* 1974; Hall 1976). In such a context, overlaps are usually interpreted as interruptions (Hafez 1991). By contrast, Egyptian and broader Arab cultures align more closely with an all-in-together model. As Nydell (2018) observes, Arab discussions are often animated, with “several people speak[ing] at once (each trying to outshout the other), interspersing their statements with gestures” (p. 93). In this context, simultaneous talk and overlaps function as signals of enthusiasm, solidarity, and engagement.

Rhythm, intonation, and volume further mark these cultural differences (Couper-Kuhlen 2001; Wennerstrom 2001). Egyptians often speak at a rapid pace, using emphatic intonation (Chahal and Hillmuth 2014; Yousri 2017) and raised volume to convey passion, sincerity, and urgency (Feghali 1997; Nydell 2018). This energetic delivery encourages quicker turn transitions and frequent overlap. In contrast, English-speaking interlocutors typically prefer slower rhythm, evenly timed speech, and falling intonation to mark turn completion (Couper-Kuhlen and Selting 1996; Wennerstrom 2001). Consequently, what Egyptians view as involvement and expressiveness may be perceived as impatience and aggression in English-speaking contexts.

An illustrative exchange, coming from Youssef’s interviews with British journalist Piers Morgan on the talk show *Piers Morgan Uncensored*, is reported below:

(1)	9:04	<i>Youssef:</i>	So, I I I told you. He’s a loser
(2)			He never kept a job
(3)			He even like failed in all of the interviews
(4)			to become like a human shield
(5)			I I I would believe
(6)	9:12	<i>Morgan:</i>	I – Let me ask you this question ↓
(7)		<i>Youssef</i>	SO, LET’S GO WITH THAT ↑ (extending his hand in front of his interlocutor, fingers spread)
(8)			NO NO NO (fast), LET’S GO, LET ME ASK YOU ↑ (pointing his left index

			towards the interlocutor)
(9)	9:16	Morgan:	{ At some point I must be able to ask you questions ↓ IF WE, IF WE ↑ It's not [?] for you Bassem ↓ IF WE AGREE, if we agree (pinching his fingers together, both hands shaking up and down) If we agree that the 14,000 casualties I mean who's counting [...] (spreading his fingers)
(10)	9:17	Youssef:	
(11)		Morgan:	
(12)	9:20	Youssef:	
(13)			
(14)			

In the exchange, Youssef frequently overlaps and interrupts (lines 6 to 12), displays animated rising intonation (lines 7 and 8), fast rhythm (line 8), high volume (lines 7, 8, 10, 12), and emphatic gestures, to secure turn-holding. While for him, these features appear to signal emotional involvement, Morgan interprets them as interruptions and “an invasion of linguistic space” (Balboni 2006, p. 27). In fact, irritated, he attempts to manage turn order (line 9), but eventually, unable to regain control, leaves Youssef to dominate the interaction. His irritation highlights a clash between Egyptian and (British) English pragmatic norms and illustrates how differing expectations regarding overlap, tempo, intonation, and volume can shape perceptions of power and politeness. In such contexts, intercultural communication requires recognising that these practices are not necessarily signs of disrespect but expressions of distinct cultural preferences for conversational flow (Scollon *et al.* 2012; Spencer-Oatey 2008).

4.2. Extra-linguistic features

4.2.1. ‘Pragmatic gestures’, interpersonal distance, and touch

Body language, often more spontaneous and less controlled than verbal expression, plays a crucial role in face-to-face interaction (Balboni 2006; Birdwhistell 1952; Hall 1976; Morris 1977). Speakers enrich or even override verbal messages through kinesics – gestures, posture, facial expressions, proxemics, and touch (Birdwhistell 1952, 1970; Martinec 2000; van Leeuwen 2004). Yet because these behaviours are culturally embedded, they are particularly vulnerable to cross-cultural misinterpretation (Balboni 2006): actions intended as humorous, emphatic, or engaging in one culture may be read as excessive, rude, or inappropriate in another.

As Samovar *et al.* (2013) and Nydell (2018) observe, Egyptian communicative style is ‘gesture-rich’ (Iverson *et al.* 2008; see also Graziano and Gulberg 2024; Kendon 2004a), characterised by marked bodily expressiveness, especially among men and less-educated speakers, whose open gestures and dynamic movements generally signal warmth, humour, and emotional engagement. This performative style is exemplified by Youssef, who often adopts a highly physical, dynamic, and heavily dramatised manner of storytelling. In his second interview with Morgan on *Talk TV*, Youssef

recounts a joke while discussing the conflict in the Middle East, accompanying almost every utterance with expansive gestures, facial mimicry, and shifts in posture, synchronically with speech (Bavelas 1990; Bavelas, Gerwing 2007) (Figure 1):

(1)	1:31:02	<i>Youssef:</i>	There was
(2)	1:31:03	<i>Morgan:</i> <i>Youssef:</i>	there it's completely wrong ↓ THERE is ↑
(3)			a hilarious (leaning towards Morgan, right hand raised with palm facing inwards, fingers slightly spread, then retreating)
(4)			documentary called <i>The Wanted 18</i> [...] (leaning in again)
(5)			So, they bought 18 cows (0.1) (moving his hand, palm facing downwards, from far to near)
(6)			18 cows (0.1) (pointing upwards with his index, thumb spread, smiling ironically)
(7)			and (0.1) they didn't know how to milk the cows
(8)			or have a cow farm [...] (rotating his torso side to side, hands raised, palms facing inwards, smiling ironically)
(9)			So they brought the cows (extending his arms forwards, then retreating them to his chest, palms facing inwards)
(10)			and they started to produce milk (clenching his fists, and moving his hands up and down, wide-eyed, smiling ironically)
(11)			and they started (moving his left arm from his chest outwards, palm facing upwards, shaking up and down)
(12)			to sell the milk to the villages (moving his arm in a semicircle, left to right, smiling ironically)
(13)			The Israeli authorities (leaning his head on his left hand)
(14)			were not very comfortable (scratching and touching his chin, gazing at his interlocutor, smiling ironically)
(15)			So one day (pointing at Morgan with his right index)
(16)			the military government came in and said
(17)			'Those cows,' and I quote (pointing up with his right index, rapidly touching his nose, smiling ironically)
(18)			"constitutes an existential threat to the national security of the state of Israel,
(19)			you need to get rid of them" (moving his left hand up and down, pointing)
(20)			And the movie goes (leaning his head backwards, head tilted back, then forwards)
(21)			about the hilarious (touching his head with both hands)
(22)			attempts of hiding those cows (shaking his arms, fingers spread, smiling ironically)
(23)			between the butchers and the houses.
(24)			And in one scene (pointing at Morgan with his right index)
(25)			a cow is actually running (extending his right arm out and upwards)
(26)			and the Israeli soldiers are running behind it (raising his hands, open-palmed, fingers claw-like, wide-eyed, smiling ironically)
(27)			And they CORNER it, and they CORNER it ↑ (leaning towards Morgan)
(28)			and they're about to KILL it ↓
(29)	1:32:09	<i>Morgan:</i> <i>Youssef:</i>	And do you know what did the cow says? ↑ (pointing at Morgan with his right finger, smiling ironically, wide-eyed)
(30)	1:32:12	<i>Youssef:</i>	(silence, both Youssef and Morgan stare at each other) (0.3)
(31)			You didn't fall for this. Cows don't speak ↓ (leaning towards Morgan, gazing at him, and shaking his hand, while smiling ironically)
(32)	1:32:14	<i>Morgan:</i>	YEEEEAAAAH ↑ (leaning backwards, head tilted back)
(33)	1:32:15	<i>Youssef:</i>	(smiling, shocked in humorous disbelief, mouth half-opened)
(34)			but you know? ↑ (pointing at Morgan with his left index)
(35)			It actually said something. ↓
(36)			You know what did it say? ↑ (leaning towards Morgan)
(37)			It said 'mooooooooooooo!!'
(38)	1:32:18	<i>Morgan:</i> <i>Youssef:</i>	but anyway! (leaning back again) Ah Ah (laughing, still in disbelief)
(39)			AH AH AH (laughing heartily, mouth wide open, head tilted back)



Figure 1
Snapshot from *Piers Morgan vs Bassem Youssef Round 2 | Two-Hour Special Interview*, showing Youssef's facial expressions, body movement, and gestures while delivering a joke. https://www.youtube.com/watch?v=rqjO5Z9Lt_M.

In this example, Youssef employs 'pragmatic gestures' (Kendon 1995, 2004a), also referred to as 'conversational gestures' (Bavelas 1990; Bavelas

and Gerwing 2007) or ‘recurrent gestures’ (Ladewig 2024), which are largely universally intelligible and shared, to animate the narrative and reinforce meaning (Kendon 2004a). Specifically, he performs iconic gestures to visualise the narrative, deictic gestures to indicate directions, and metaphoric gestures to represent abstract ideas (McNeill 1992, 2005; Norris 2004a). For instance, he enacts the milking motions (line 10 and Figure 1, fifth scene), sweeping arm gestures to describe the distribution of milk (line 12 and Figure 1, sixth scene), and frantic movements to mimic hiding (line 22 and Figure 1, seventh scene) and chase (line 25 and Figure 1, eighth scene). Such ‘representative actions’ (Martinec 2000) visually illustrate the story, amplifying its comedic and dramatic tones and heightening its absurdity. Beat gestures (McNeill 1992, 2005), or ‘indexical actions’ (Martinec 2000) (see line 19), further punctuate his speech, reinforcing rhythm and emphasis. This animated bodily delivery may sometimes be perceived as entertaining, charismatic, and persuasive, even when culturally incongruent. However, it can also appear excessive and impolite to an interlocutor like Morgan, from a ‘gesture-poor’ culture (Iverson *et al.* 2008; see also Graziano and Gulberg 2024; Kendon 2004a) that values restraint, and where movement intensity may be misread as aggression, mockery, or loss of self-control (Hall 1976; Knapp *et al.* 2014; Nydell 2018; Remland 2009).

In addition to gestures, Youssef frequently leans forward, narrowing the physical distance between himself and Morgan, occasionally resorting to touch. According to Hall (1959), proxemics may be sociofugal (discouraging closeness) or sociopetal (promoting closeness). While British and other anglophone cultures typically adopt a sociofugal orientation, requiring greater interpersonal distance, often five feet or more (Feghali 1997), Egyptians tend towards sociopetal arrangements (Hall 1959; Nydell 2018), feeling comfortable at two feet or less (Feghali 1997; Balboni 2006), and being even contact-oriented (Busà *et al.* 2023). Consequently, proximity and touch that signal friendliness in Egyptian culture may be perceived as an aggressive and impolite invasion of personal space (Feghali 1997) in British contexts. This cultural divergence is evident at the climax of Youssef’s joke (lines 28-39), when he suddenly leans in and shakes Morgan’s hand (line 30). While intended as a sign of involvement and solidarity (Hall 1959; Nydell 2018), Morgan’s stiff, motionless response betrays discomfort. His unease is even clearer in another moment of the interview (minutes 1:16:24-1:16:27), where he subtly leans his head back as Youssef intrusively moves closer, giving the impression of feeling annoyed or even threatened (Figure 2).



Figure 2

Snapshot from *Piers Morgan vs Bassem Youssef Round 2 | Two-Hour Special Interview*, showing Youssef leaning towards the interviewer, who slightly retreats.

https://www.youtube.com/watch?v=rqjO5Z9Lt_M.

4.2.2. Facial expression and gaze

The face – the eyebrows, mouth, and gaze direction in particular (Holler 2025) – constitutes one of the most powerful communicative channels, often conveying emotions (Ekman and Friesen 1971; Haller and Rambaud 1987), “communicating intentions and [...] facilitat[ing] pragmatically appropriate responding” (Holler 2025, p. 750).

However, facial signals are culturally specific (Holler 2025) and may lead to misinterpretations in intercultural encounters. In Youssef’s performance (Figure 2), shifts in eyebrow position, ironic smiles, and bursts of open laughter accompany his narrative, reinforcing its playful and satirical tone. Yet this very intensity risks clashing with the British interviewer’s expectations of politeness. In a British communicative framework, indeed, exaggerated expressions and loud laughter may appear overwhelming, placing the interlocutor in a passive role or even an uncomfortable position. Morgan’s restrained responses, limited to subtle eyebrow raises and a half-smile with the corners of his mouth drawn back (Figure 1, tenth scene), signal this contrast. His minimal expressiveness illustrates polite acknowledgement but also suggests unease, mock surprise, or failure to fully share Youssef’s humour. This juxtaposition not only underscores cultural divergence in facial expressiveness but also reveals moments of nonverbal misalignment that may generate embarrassment or impressions of impoliteness.

Eye behaviour, or oculusics (Andersen 2010), provides another key layer of interaction. Gaze regulates turn-taking, offers feedback, and conveys emotions (Argyle and Cool 1976; Ekman and Friesen 1971; Goodwin 1981; Kendon 1967). While universal in functions, its use is strongly culture-specific (Busà *et al.* 2023). For example, Egyptians, like other Arabs, often sustain prolonged eye contact as a sign of sincerity and attentiveness (Balboni 2006). However, outsiders may perceive such intensity as intrusive staring or even as a claim to dominance (Samovar *et al.* 2013). Such mismatches may

cause discomfort or even conflict in intercultural exchanges. In the second interview with Morgan, Youssef repeatedly fixes his gaze on his interlocutor for extended periods, sometimes lasting over ten seconds. In one striking instance (minutes 32:31-32:4), he maintains direct eye contact with widened eyes (Figure 3), producing a stare that may be read as emphatic and engaging in his cultural frame, but potentially unsettling in Morgan's.



Figure 3

Snapshot from *Piers Morgan vs Bassem Youssef Round 2 | Two-Hour Special Interview*, showing Youssef staring at his interlocutor, wide-eyed.

https://www.youtube.com/watch?v=rqjO5Z9Lt_M.

4.2.3. A focus on 'emblematic gestures'

As seen in Section 4.2.1, Youssef employs a wide repertoire of gestures – pointing, mimicking, special framing – to animate his arguments. Beyond these 'pragmatic gestures' (Kendon 1995, 2004a), he frequently performs 'emblematic gestures' (Kendon 1995, 2004a), whose meanings are conventional and culturally agreed-upon (Abner *et al.* 2015; Archer 1997; Birdwhistell 1970; Kecskés 2014; Kendon 1995, 2004a; Nydell 2018; Samovar *et. al* 2013), differing drastically across cultures (Cooperrider 2020). As a result, these gestures often generate misconstructions, misinterpretations, and even unintended offense.

The purse-hand gesture (Kendon 1995) is perhaps the most recognisable and emblematic among Egyptian gestures. In this movement, fingers are pinched together and the hand(s) move(s) slowly up and down (Figure 4). It usually signals "wait", "slow down", or "relax", though, depending on speed and accompanying expression, it can also convey impatience or a veiled threat. In an interview with the Canadian-American *The Nelk Boys*, Youssef performs this gesture repeatedly, adapting its meaning to context.

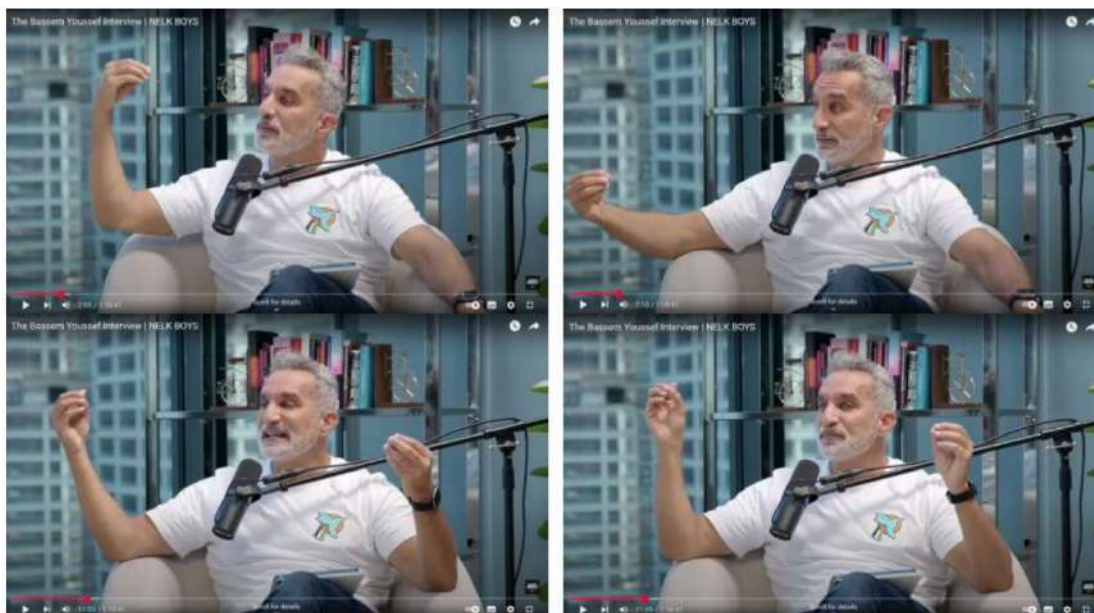


Figure 4
Snapshot from *The Bassem Youssef Interview | NELK BOYS*, showing Youssef making the purse-hand gesture.

<https://www.youtube.com/watch?v=GW-yaLz-3OE>.

At times, this ‘indexical action’ (Martinec 2000) expresses questioning and pleading for understanding. For example, in the first two scenes in Figure 4, Youssef performs this gesture while uttering the following words:

(1)	6:50	<i>Youssef:</i>	You’re thirty years old and you need to be aware (extending his right hand towards the two interlocutors, fingers cupped together)
(2)			of what’s happening in the world (raising his left hand, fingers cupped, both hands swinging up and down, eyebrow raised)
(3)			and that you reach
(4)			and your platforms mean something
(5)			and it affects people.
(6)			And I think the biggest reason (rotating palms downwards, hands swinging up and down)
(7)			why people are so upset for you,
(8)			not because you brought someone (spreading his arms outwards to left and right, fingers still cupped together)
(9)			And you didn’t contest him
(10)			But that it is the talking point that he used (relaxing his left arm, swinging his right hand with fingers still cupped together)
(11)			is the same talking points that have been used for two years
(12)			to kill thousands of people. Right?

The tag question at the end (line 12) and the raised eyebrow (line 2 and Figure 4, second scene) confirm and reinforce the questioning nature of Youssef’s discourse and his plea for understanding. At other times, the same gesture conveys frustration and irritation, as in his first interview with Morgan (see example in Section 4.1.1, line 12), where, exasperated by interruptions, he uses it while struggling to retain the floor. In other contexts, the gesture carries moral weight. During the *Nelk Boys* interview (Figure 4,

third and fourth scenes), Youssef employs it to underline the absurdity of racist tropes:

(1)	10:52	<i>Youssef:</i>	I don't know who edited
(2)			that.. if..
			it seems that
			Mousad edited it
(3)			but if you guys edited,
(4)			that is fucked up (raising his left hand, palm facing downwards, swinging up and down)
(5)			because it (rotating his left hand, palm facing inwards)
			just.. j.. just
(7)			like it.. (spreading his fingers)
(8)			solidifies (raising his right hand cupped, hands slowly swinging up and down, the corners of his mouth drawn down)
(9)			the same racist tropes
(10)			about a group of people
(11)			that being slaughtered and starved to death right now
(12)			and then, they distract you with questions like (.) (spreading his fingers again, hands semi-cupped)
(13)			it is Hamas! (spreading arms and hands outwards to left and right, fingers widely spread)

Here, the movement highlights the gravity of his words, pointing to justice and challenging the interlocutor's (or the editors') perspective. Its force is amplified by Youssef's facial expression, featuring the corners of his mouth drawn down in visible irritation (line 8, and Figure 4, third scene). The gesture can also take on lighter tones. At the *India Today Conclave 2024*, for instance, Youssef uses it when saying "I'm very very pleased to be here!" to emphasise enthusiasm and amplify the meaning of words. However, across these examples, since the gesture is unfamiliar to Canadian-American or Indian contexts, it fails to carry its intended weight, losing both semantic resonance and pragmatic effects.

A second local Egyptian hand gesture is the 'what's going on?', made by twisting the wrist clockwise and anticlockwise with the palm facing down. It generally conveys confusion, outrage, or concern. During his speech at *India Today Conclave 2024*, Youssef uses this gesture (Figure 5) to denounce Israeli propaganda during the war in Gaza:

(1)	12:33	<i>Youssef:</i>	They did it through lies (pointing downwards with his left index)
(2)			propaganda (twisting his right wrist clockwise and anticlockwise)
(3)			twisting facts
(4)			and emotional blackmail (twisting his left wrist clockwise and anticlockwise)



Figure 5

Snapshot from *India Today Conclave 2024: Viral from Israel-Gaza War Stinging Comic Commentary by Egyptian Comedian*, showing Youssef performing a twisting-hand gesture. https://www.youtube.com/watch?v=_04qco8OjkY.

However, in anglophone contexts, including India, this gesture's meaning is opaque, risking pragmatic failure; like the 'be patient' movement, it may be read as a mere filler rather than an appeal for moral urgency.

Some gestures share identical hand movements across cultures but carry divergent meanings, making them particularly prone to confusion in intercultural contexts. One instance of a hand gesture that is not unique to Egyptian culture, yet has an exclusive meaning in Egypt, is the emphatic 'I swear to God', where the thumb and forefinger form a circle. In Western cultures, it is positively connotated, signalling "okay"; in Egypt, instead, it carries hostile, sometimes threatening undertones, often intensified with a baring of teeth (Samovar *et al.* 2013). This gesture recurs frequently in Youssef's speeches, as shown in the snapshot from the *Piers Morgan vs Bassem Youssef Round 2* video (Figure 6). In this instance, it accompanies his reference to the Holocaust, presented as the most horrible genocide in modern history:

(1)	46:10	<i>Youssef:</i>	As you see (extending his arms towards his interlocutor, palms facing inwards, teeth bared)
(2)			the Holocaust happened (retreating his arms)
(3)			the most
(4)			orchestrated industrialised (joining his thumb and forefinger in a circle, other fingers extended)
(5)			horrible genocide (moving his left arm up and down, fingers still connected)
(6)			in our modern time (spreading his fingers, palm facing downwards)



Figure 6

Snapshot from *Piers Morgan vs Bassem Youssef Round 2 | Two-Hour Special Interview*, showing Youssef with bared teeth, performing a “I swear to God” threatening gesture.

https://www.youtube.com/watch?v=rqjO5Z9Lt_M.

If isolated from facial expression and verbal context, the gesture’s meaning risks being inverted; what Youssef intends as anger or disapproval may be read by the British journalist as reassurance, consent, or positivity, heightening the chance of misinterpretation.

Similarly, the rotating index finger near the temple in Egypt means ‘think about it!’ or “use your head”, but in many Western English-speaking contexts, it is read as “are you crazy?”, an almost opposite meaning. In one instance from the same *Piers Morgan vs Bassem Youssef Round 2* video, Youssef uses this metaphorical gesture (Figure 7), while citing the reflections of an Israeli historian, arguing that many Israelis lack a real concept of peace due to the persistent war in their territory:

(1)	58:16	<i>Youssef:</i>	{ There is a boo – (extending his right arm, fingers cupped) I know you would say most of his cabinet The.. the.. (fast)
(2)	58:17	<i>Morgan:</i>	
(3)	58:19	<i>Youssef:</i>	
(4)			there is a book called <i>The Fear of Peace</i> , (touching his chin)
(5)			it’s called by Moshe Zimmermann (holding his throat area)
(6)			an Israeli historian (extending his right arm towards his interlocutor, palm facing downwards)
(7)			and he said like (slightly retreating his hand, fingers claw-like)
(8)			the average Israeli citizen (swinging his hand up and down)
(9)			does not have a vision of Peace
(10)			because for seventy years this is a country (shaking his head side to side)
(11)			that has been the military (.)
(12)			war has been going on for long [...] (raising his right hand, moving it back and forward, fingers held close together. Gazing at his interlocutor, wide-eyed, and smiling ironically)
(13)			So the whole idea of peace (rotating his right index finger near his temple)
(14)			is not even there (slightly getting up from the chair, still rotating his finger)



Figure 7
Snapshot from *Piers Morgan vs Bassem Youssef Round 2 | Two-Hour Special Interview*,
showing Youssef performing a ‘thinking’ gesture.
https://www.youtube.com/watch?v=rqjO5Z9Lt_M.

In this instance, Youssef associates the movement with the act of thinking or envisioning. However, while Youssef uses it in a descriptive, non-judgmental way, the British interlocutor may misread it as highlighting the absurdity of the idea.

Notably, in other moments of the same video, Youssef resorts to the Egyptian equivalent of ‘are you crazy?’ by pointing both index fingers to the sides of his head. While in anglophone cultures, this same configuration is used to convey that a person is thinking hard and using their brain, emphasising intelligence and concentration, in contrast, in Egypt, it is often used in debates or animated conversations to urge interlocutors to think logically, recall something important, or reconsider facts and their position carefully. In the instance in Figure 8, Youssef performs this gesture to mark his interlocutor’s reasoning as irrational and foolish, accompanying his words that denounce the unreasonable act of killing Israelis:

(1)	1:21:28	<i>Youssef:</i>	The only two crutches that (rotating his hands, index and thumb pinched, palms facing inwards)
(2)			that that is is like Abdullah (extending his arm to the right, palm facing downward, fingers spread).
(3)			and Hamas (raising both hands at chest level, fingers spread and slightly curved downwards)
(4)			A few militant
(5)			Thousands of militant (rotating his palms upwards, fingers spread)
(6)			is is that really formed an existential (rotating his palms downwards again, fingers spread and slightly curved)
(7)			like a threat especially (relaxing his right arm, raising his left hand to his temple, fingers touching the side of his head, brows furrowed)

- (8) If I know that over 13 years (pointing and pressing both index fingers to temples, furrowed brows)
- (9) Only.. not not.. 69 Israelis were killed?



Figure 8

Snapshot from *Piers Morgan vs Bassem Youssef Round 2 | Two-Hour Special Interview* showing Youssef performing the ‘are you crazy?’ gesture.

https://www.youtube.com/watch?v=rqjO5Z9Lt_M.

This message is reinforced by his facial expression, which reveals indignation. His eyebrows are drawn inward and downward, his eyes are slightly narrowed, conveying scrutiny, and his lips are pressed together and slightly pulled down at the corners to signal strong disapproval (Figure 8). However, without accompanying verbal or facial cues, the British interlocutor might interpret the gesture differently, associating it with deep thought and positivity rather than disapproval. These examples illustrate how intercultural encounters are sites of indexical reversibility, where identical forms can convey opposite meanings across boundaries (Kendon 2004a), meanwhile corroborating Holler’s (2025) idea that facial signals “significantly improve participants’ accuracy” (p. 753) in distinguishing their pragmatic intentions.

Yet, while these gestures may produce ambiguous meanings and potential misinterpretations, others even carry the risk of face-threatening acts (Brown and Levinson 1987), potentially causing offense (Spencer-Oatey 2008). For example, in his interactions with Morgan, the American podcaster Lex Friedman and the British-American broadcaster and writer Mehdi Hasan, Youssef frequently touches his chin, face, or neck (see example in Section 4.2.1., lines 13 and 21), caresses and scratches his beard, leans his head on his hand (Figure 9), or points at his interlocutor(s) (Figure 10).



Figure 9

Snapshot from *Bassem Youssef: Israel-Palestine, Gaza, Hamas, Middle East, Satire & Fame | Lex Fridman Podcast #424*, showing Youssef resting his head on his hand during the debate. <https://www.youtube.com/watch?v=sG8u6owzad4>.

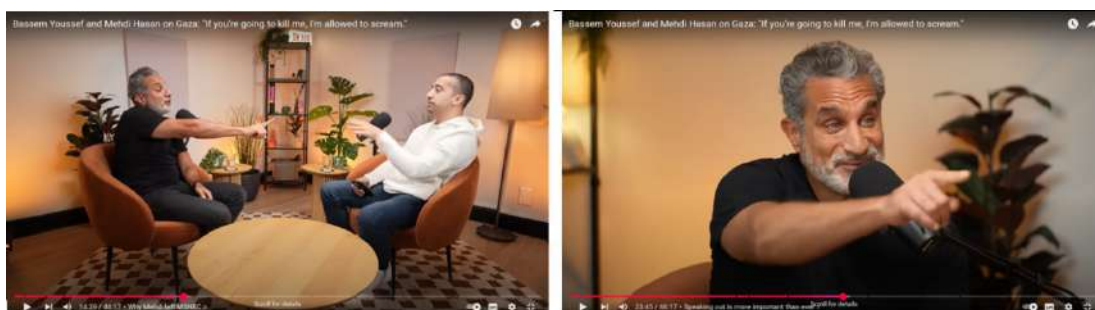


Figure 10

Snapshot from *Bassem Youssef and Mehdi Hasan on Gaza. "If you're going to kill me, I'm allowed to scream."*, showing Youssef pointing at his interlocutor. <https://www.youtube.com/watch?v=LWmL1HfEs>.

In the Egyptian communicative context, these gestures commonly signal stress, concern, reflective attention, or evaluative thought. However, British or Anglo-American interlocutors may interpret them as signs of distraction, disinterest, and boredom, giving rise to perceptions of impoliteness and disrespect. Similarly, Youssef's pointing gestures, intended to direct attention and coordinate interaction (Kendon 2004b), may be perceived as aggressive or accusatory, creating the impression of personal attack.

5. Discussion

By observing Youssef's performances in English-mediated discourses on YouTube, the present analysis offers empirical insight into culturally embedded Egyptian nonverbal practices in cross-cultural contexts. Specifically, it shows that Youssef's L1 multimodal features continue framing meaning and shaping interaction in anglophone contexts, potentially leading to pragmatic variability. The transfer of culturally embedded

multimodal norms sometimes seems to enhance communication, adding vividness to the interaction, but other times, it appears to clash with interlocutors' expectations, leading to misunderstanding, unintended impoliteness, or even offense (Busà *et al.* 2023; Hall 1959; Scollon *et al.* 2012; Spencer-Oatey 2008). This supports Kecskés' (2014) model of intercultural pragmatics, which views communication as a negotiation between L1 and L2 (or the foreign language) pragmatic systems rather than a complete shift from one to the other. Youssef's communicative repertoire – including intrusive overlap, emphatic volume, fast rhythm, animated gestures, dynamic body movement, short interpersonal distance, touch, communicative facial expressions, and strong and prolonged eye contact – seems to reflect Egyptian cultural values of sociability, immediacy, warmth, and involvement, especially under emotional engagement (Samovar *et al.* 2013; Nydell 2018). However, while common in Egyptian contexts (Nydell 2018), these norms may be perceived as excessive, assertive, or inappropriate when filtered through more restrained and sociofugal conventions (Hall 1959; Feghali 1997), potentially generating discomfort, embarrassment, and a perception of impoliteness. A major focus in the analysis is on 'emblematic gestures' (Kendon 1995, 2004a) and on those that, while formally identical, may be decoded differently across cultures. Findings indicate that such cases, including the purse-hand or twisting-wrist movement, do not guarantee cross-cultural intelligibility. They may be meaningless in anglophone contexts, produce divergent connotations, or sometimes even develop unintended or hostile undertones, resulting in pragmatic breakdowns, confusion, or inadvertent face-threatening acts (Brown and Levinson 1987).

Overall, these findings reaffirm the cultural relativity of nonverbal systems of meaning and support the idea that chronemics, vocalics, kinesics, proxemics, oculusics, and haptics are not universal but culturally encoded and decoded differently across contexts (Andersen 2010).

6. Conclusion

Through the illustrative case of the Egyptian actor Bassem Youssef's multimodal communicative style in anglophone settings, this study emphasises the cultural specificity of prosodic and nonverbal behaviours and their potential impact on intercultural communication outcomes. It suggests that when these behaviours are not shared, carry different meanings across cultures, or occur between speakers with differing communicative styles (e.g., gesture-rich vs gesture-poor), pragmatic failure may arise. The findings thus would confirm that effective intercultural communication relies not only on linguistic competence but also on the ability to interpret and manage multimodal signals.

By doing so, this research contributes to the study of multimodal interaction in intercultural communication, affirming the value of integrating multimodality and intercultural pragmatics approaches. Beyond its descriptive contribution, the study also highlights an important intercultural implication. In line with Gudykunst (2004) and Holliday (2011), it emphasises that effective communication requires sensitivity to how multimodal cues are produced and received across cultures. It emphasises the need to recognise that what appears as assertiveness, intrusion, or excessive expressivity in one culture may function as warmth, immediacy, or solidarity in another. Cultural awareness of such differences, therefore, becomes important for interpreting multimodal signals and can reduce negative stereotyping and misattribution of intention, fostering empathy and greater tolerance towards communicative diversity. In this sense, this study contributes to the broader discussion of intercultural competence.

Nevertheless, although providing valuable insights, this study has some limitations. First, it is constrained to the single case of Youssef. Second, although overly formal interviews were excluded, the analysed discourses still stem from settings where both language and nonverbal behaviour are to some extent regulated. Additionally, the dataset is drawn from interviews with a comedian, which are inevitably performative. To address these limitations, future research should broaden the corpus by including a wider range of Egyptian speakers, also considering the internal socio-cultural diversity of the population. It should also extend the analysis to more spontaneous and informal interactions beyond interviews with Egyptian public figures. Such work would offer deeper insights into how embodied Egyptian norms are negotiated in everyday intercultural encounters.

Bionote: Lucia La Causa holds a PhD in Sciences of Interpretation. She is currently a Research Fellow and an Adjunct Professor at the University of Catania. Her research interests focus on English sociolinguistics, World Englishes, intercultural pragmatics, contact literature, audiovisual discourses and translation, and CDA. She has published in national and international journals and edited volumes. Her latest works include: *Existing World Englishes Models and Beyond, Filling The 'Grey Area'* (Loffredo Editore, 2025), “Autobiographical narrations of sexual harassment on online blogs as tools for Egyptian women’s empowerment: a linguistic and critical analysis” (*European Scientific Journal*, 2025), and “Censorship and identity in Arabic subtitling of American films” (*I-Land Journal*, 2024).

Author’s address: lucia.lacausa@unict.it

References

- Abner N., Cooperrider K. and Goldin-Meadow S. 2015, *Gesture for Linguists: A Handy Primer*, in “Language and Linguistics Compass” 9 [11], pp. 437-451.
- Abu-Arqoub I.A. and Alserhan F.A. 2019, *Nonverbal Barriers to Effective Intercultural Communication*, in “Utopía y Praxis Latinoamericana” 24 [5], pp. 307-316.
- Adly M. 2021, *Multimodal Investigation of Egyptian Politeness in Relation to Gender, Social Distance, and Power*, in “Egyptian Journal of Linguistics and Translation” 5 [1], pp. 1-29.
- Ali Aldhahri S. 2024, *A Comparative Study of Turn-taking in Arabic and English Conversations: Implications for Language Learners*, in “International Journal for Scientific Research (IJSR)” 3 [10], pp. 106-122.
- Andersen P.A. 2010, *Nonverbal Communication*, in Samovar L.A., Porter R.E. and McDaniel E.R. (ed.), *Intercultural Communication: A Reader*, 13th edn., Wadsworth, Boston, pp. 293-312.
- Archer D. 1997, *Unspoken Diversity: Cultural Differences in Gestures*, in “Qualitative Sociology” 20, pp. 79-105.
- Argyle M. and Cook M. 1976, *Gaze and Mutual Gaze*. Cambridge University Press, Cambridge.
- Balboni P.E. 2006, *Intercultural Communicative Competence: A Model*, Guerra Edizioni, Perugia.
- Balboni P.E. 2015, *La comunicazione interculturale e l'approccio comunicativo: dall'idea allo strumento*, in “Educazione Linguistica – Language Education (EL.LE)” 4 [1], pp.1-20.
- Bavelas J.B. 1990, *Nonverbal and Social Aspects of Discourse in Face-to-face Interaction*, in “Text” 10, pp. 5-8.
- Bavelas J.B. and Chovil N. 2006, *Nonverbal and Verbal Communication. Hand Gestures and Facial Displays as Part of Language Use in Face-to-face Dialogue*, in Manusov V. and Patterson M.L. (ed.), *The Sage Handbook of Nonverbal Communication*, Sage Publications, Thousand Oaks (CA), pp. 97-115.
- Bavelas J.B. and Gerwing J. 2007, *Conversational Hand Gestures and Facial Displays in Face-to-face Dialogue*, in Fiedler K. (ed.), *Social Communication*, Psychology Press, London, pp. 283-308.
- Beltrán-Planques V. and Querol-Julián M. 2018, *English Language Learners' Spoken Interaction: What a Multimodal Perspective Reveals about Pragmatic Competence*, in “System” 77, pp. 80-90.
- Birdwhistell R. 1952, *Introduction to Kinesics: (An Annotation System for Analysis of Body Motion and Gesture)*, Department of State Foreign Service Institute, Washington, D.C.
- Birdwhistell R. 1970, *Kinesics and Context: Essays on Body-motion Communication*, University of Pennsylvania Press, Philadelphia (PA).
- Brown P. and Levinson S.C. 1987, *Politeness: Some Universals in Language Usage*, Cambridge University Press, Cambridge.
- Busà M.G., Facciani C. and Notaro A. 2023, *Multilingual Communication: The Role of Gaze, Physical Contact, and Time Perception and Organization in Intercultural Interactions*, in “De Europa” 6 [1], pp. 103-122.
- Chahal D. and Hllmuth S. 2014, *The Intonation of Lebanese and Egyptian Arabic*, in Jun S.A. (ed.), *Prosodic Typology II*, Oxford University Press, Oxford, pp. 365-404.
- Cooperrider K. 2020, *Universal and Diversity in Gesture: Research Past, Present, and*

- Future*, in “Gesture” 18 [2-3], pp. 209-238.
- Couper-Kuhlen E. 2001, *Interactional Prosody: High Onsets in Reason-for-the-call Turns*, in “Language in Society” 30 [1], pp. 29-53.
- Couper-Kuhlen E. and Selting M. 1996, *Prosody in Conversation: Interactional Studies*, Cambridge University Press, Cambridge.
- Ekman P. and Friesen W.V. 1971, *Constants across Cultures in the Face and Emotion*, in “Journal of Personality and Social Psychology” 17 [2], pp. 124-129.
- Eladawy M.A. 2021, *The Role of Non-linguistic and Kinesics in Daily Life Communications in Egypt*, in “Social’nojeekonomicheskoe upravlenie: teorija i praktika” 17 [4], pp. 76-81.
- Feghali E. 1997, *Arab Cultural Communication Patterns*, in “Intercultural Journal of Intercultural Relations”, 21 [3], pp. 345-378.
- Goodwin C. 1981, *Conversational Organisation: Interaction between Speakers and Hearer*, Academic Press, New York.
- Graziano P. and Gullberg M. 2024, *Gestural Communication across Cultures: The Role of Gestures in Social Interaction*, in “European Journal of Linguistics” 3 [1], pp. 25-37.
- Gumperz J.J. [1982] 2009, *Discourse Strategies*, Cambridge University Press, Cambridge.
- Gudykunst W.B. 2004, *Bridging Differences: Effective Intergroup Communication*, 4th ed., Sage Publications, Thousand Oaks (CA).
- Hafez O.M. 1991, *Turn-taking in Egyptian Arabic: Spontaneous Speech vs. Drama Dialogue*, in “Journal of Pragmatics” 15, pp. 59-81.
- Hall E.T. 1959, *The Silent Language*, Doubleday, New York.
- Hall E.T. 1966, *The Hidden Dimension*, Doubleday, New York.
- Hall E.T. 1974, *Handbook for Proxemic Research*, Society for the Anthropology of Visual Communication, Washington, D.C.
- Hall E.T. 1976, *Beyond Culture*, Doubleday, New York.
- Haller B. and Rambaud P. 1987, *Le visage parle*, Balland, Paris.
- Handriani P., Dinata A.P., Siregar F.A. and Siregar D.Y. 2025, *Politeness Strategies in Cross-cultural Communication: A Pragmatic Approach*, in “Semantik: Jurnal Riset Ilmu Pendidikan, Bahasa dan Budaya” 3 [1], pp. 126-132.
- Haryanti P., Saddhono K. and Anindyarini A. 2023, *Multimodal as a New Perspective in Pragmatics in the Digital Era: Literature Review*, in *Proceeding of the the 3rd International Conference of Humanities and Social Science (ICHSS)*, pp. 494-501.
- Holler J. 2025, *Facial Clues to Conversational Intentions*, in “Trends in Cognitive Sciences” 29 [8], pp. 750-762.
- Holliday A. 2011, *Intercultural Communication and Ideology*, Sage Publications, Thousand Oaks (CA).
- Indarti D. 2024, *Multimodalities and Conversational Implicature in Cross-Cultural Pragmatic Context: A Systematic Review*, in “International Journal of Social Science and Human Research” 7 [12], pp. 9271-9280.
- Ishihara N. 2018, *Intercultural Pragmatic Failure: Sociocultural Aspects of English Language Teaching Intercultural Competence*, in *The TESOL Encyclopedia of English Language Teaching*. <https://onlinelibrary.wiley.com/doi/abs/10.1002/9781118784235.eelt0284> (15.11.2025).
- Iverson J.M., Goldin-Meadow S. and Reynolds M. 2008, *The Role of Gesture in Early Language Development: Gesture-rich and Gesture-poor Cultures*, in “Developmental psychology” 44 [5], pp. 1248-1263.
- Kecskés I. (ed.) 2022, *The Cambridge Handbook of Intercultural Pragmatics*, Cambridge University Press, Cambridge.
- Kecskés I. 2014, *Intercultural Pragmatics*, Oxford University Press, Oxford.

- Kendon A. 1967, *Some Functions of Gaze-direction in Social Interaction*, in “Acta Psychologica” 26, pp. 22-63.
- Kendon A. 1995, *Gestures as Illocutionary and Discourse Structure Makers in Southern Italian Conversation*, in “Journal of Pragmatics” 23 [3], pp. 247-279.
- Kendon A. 2004a, *Gesture: Visible Action as Utterance*, Cambridge University Press, Cambridge.
- Kendon A. 2004b, *On Pointing*, in Kendon A. (ed.), *Gesture. Visible Action as Utterance*, Cambridge University Press, Cambridge, pp. 199-224.
- Knapp M.L., Hall J.A. and Horgan T.G. 2014, *Nonverbal Communication in Human Interaction*, 8th ed., Cengage, Boston.
- Kress G. and van Leeuwen T. 2001, *Multimodal Discourse: The Modes and Media of Contemporary Communication*, Arnold, London.
- Ladewig S.H. 2024, *Recurrent Gestures and Their Role in Communication*, in Müller C., Cienki A.C., Fricke E., Ladewig S.H., McNeill D. and Bressem J. (ed.), *Body-Language-Communication: An International Handbook on Multimodality in Human Interaction*, De Gruyter Mouton, Berlin/Boston, pp. 1558- 1575.
- Lopriore L. and Grazzi E. 2016, *Intercultural Communication: New Perspectives from ELF*, Roma TrE-Press, Roma.
- Martinec R. 2000, *Types of Processes in Action*, in “Semiotica” 130 [3-4], pp. 243-268.
- McNeill D. 1992, *Hand and Mind: What Gestures Reveal about Thought*, University of Chicago Press, Chicago (IL).
- McNeill D. 2005, *Gesture and Thought*, University of Chicago Press, Chicago (IL).
- Morris D. 1977, *Manwatching*, Abrams, New York.
- Mühleisen S. 2011, *Pragmatics in Multilingual Language Situations*, in Andersen G. and Aijmer K. (ed.), *Pragmatics of Society*, Mouton de Gruyter, Berlin, pp. 115-134.
- Norris S. 2004a, *Analyzing Multimodal Interaction: A Methodological Framework*, Routledge, New York/London.
- Norris S. 2004b, *Multimodal Discourse Analysis: A Conceptual Framework*, in Levine P. and Scollon R. (ed.), *Discourse and Technology: Multimodal Discourse Analysis*, Georgetown University Press, Washington, D.C., pp. 101-115.
- Norris S. 2011, *Identity in (Inter)action: Introducing Multimodal Interaction Analysis*, Mouton de Gruyter, Berlin.
- Norris S. 2019, *Systematically Working with Multimodal Data: Research Methods in Multimodal Discourse Analysis*, Wiley-Blackwell, Hoboken (NJ).
- Norris S. 2020, *Multimodal theory and Methodology: For the Analysis of (Inter)action and Identity*, Routledge, London.
- Norris S. 2021, *Multimodal (Inter)action Analysis in a Nutshell: Philosophy, Theory, Method and Methodology*, in “Qualitative Video Research Reports” 2, pp. 1-7.
- Nydell M. 2018, *Understanding Arabs. A Contemporary Guide to Arab Society*, 6th ed., Intercultural Press, Los Angeles (CA).
- O'Halloran K.L., Tan, S. and Marissa, K.L.E. 2014, *Multimodal Pragmatics*, in Schneider K.P. and Barron A. (ed.), *Pragmatics of Discourse*, Mouton de Gruyter, Berlin, pp. 239-268.
- Payratò L. 2017, *Pragmatics and Multimodality. A Reflection on Multimodal Pragmatics*, in Giora R. and Haugh M. (ed.), *Doing Pragmatics Interculturally: Cognitive, Philosophical, and Sociopragmatic Perspectives*, Mouton de Gruyter, Berlin, pp. 293-312.
- Raddawi R. (ed.) 2015, *Intercultural Communication with Arabs. Studies in Educational, Professional and Societal Contexts*, Springer, Singapore.
- Remland M.S. 2009, *Nonverbal Communication in Everyday Life*, 3rd ed., Allyn & Bacon,

Boston.

- Sacks H., Schegloff E.A. and Jefferson G. 1974, *A Simplest Systematics for the Organization of Turn-Taking for Conversation*, in "Language" 50 [4], pp. 696-735.
- Samovar L.A., Porter R.E., McDaniel E.R. and Roy C.S. 2013, *Communication between Cultures*, 8th ed., Wadsworth, Boston.
- Schroeder M.L. 2017, *A Research Framework for Second Nonverbal Code Acquisition*, in "Globe: A Journal of Language, Culture and Communication" 5, pp. 102-134.
- Scollon R., Scollon S.W. and Jones R.H. 2012, *Intercultural Communication: A Discourse Approach*, 3rd ed., Wiley-Blackwell, Malden (MA).
- Spencer-Oatey H. 2008, *Culturally Speaking: Culture, Communication and Politeness Theory*, Continuum, London.
- Stivers T., Enfield N.J., Brown P., Englert C., Hayashi M., Heinemann T., Hoymann G., Rossano F., de Ruiter J.P., Yoon K-E and Levison S.C. 2009, *Universals and Cultural Variation in Turn-Taking in Conversation*, in "Proceeding of the National Academy of Science USA" 106 [26], pp. 10587-10592.
- Tannen D. 2005 [1984], *Conversational Style*, Oxford University Press, Oxford.
- Ting-Toomey S. 1999, *Communicating across Cultures*, Guilford Press, New York.
- van Leeuwen T. 2004, *Ten Reasons Why Linguists Should Pay Attention to Visual Communication*, in Levine P. and Scollon R. (ed.), *Discourse and Technology. Multimodal Discourse Analysis*, Georgetown University Press, Washington, D.C., pp. 7-19.
- Wennerstrom A. 2001, *The Music of Everyday Speech: Prosody and Discourse Analysis*, Oxford University Press, Oxford.
- Yousri E.M. 2017, *Perception of English Intonation by Egyptians*, in "Social'nojeekonomicheskoe upravlenie: Teoriya i praktika" 17 [4], pp. 76-81.
- Zaharna R.S. 1995, *Understanding Cultural Preferences of Arab Communication Patterns*, in "Public Relations Review" 21 [3], pp. 241-255.

SINCERE OR OPPORTUNISTIC CORPORATE INCLUSION? A Pragmatic and Multimodal Corpus-informed Approach to LGBTQ+ (In/Ex)clusion in CSR and ESG Disclosures

FEDERICO ZAUPA
UNIVERSITY OF MODENA AND REGGIO EMILIA

Abstract – In corporate social responsibility (CSR) and environmental, social, and governance (ESG) communication on diversity, equity, and inclusion, companies have increasingly addressed the inclusion and support of LGBTQ+ people and communities. Given the relevance of this topic and the absence of linguistic, pragmatic, and multimodal studies examining whether such communication is sincere or opportunistic, this research analysed a corpus of CSR and ESG disclosures published between 2020 and 2023 by U.S.-based international companies. The study adopted a multimodal approach to investigate corporate communication on LGBTQ+ inclusion by identifying lexico-phraseological patterns, linguistic choices signalling adherence to conversational maxims and conveying transparency or vagueness, and the interplay between these linguistic features and the visual patterns emerging from the integrated images. Findings suggest a communicative focus on corporate practices aimed at ensuring equality for LGBTQ+ employees and improving the social well-being of local LGBTQ+ communities in the areas where companies operate. While such information seems to be disclosed transparently – i.e., complete, accurate, and relevant – the simultaneous use of vague lexical choices undermines the degree of information transparency. The visual analysis revealed, on the one hand, visibility of specific sexual and gender identities; on the other hand, representations of people in heterogeneous groups challenge diversity narratives and reinforce a holistic and not an identity-specific approach to LGBTQ+ inclusion. Overall, although the seemingly factual communication and the visual attention to diverse identities may signal sincere corporate engagement, linguistic vagueness points to a communication strategy that is, at least in part, constructed and potentially opportunistic.

Keywords: CSR; LGBTQ+ discourses; pragmatics; multimodality; sincerity vs. opportunism.

1. Introduction

In recent years, *diversity, equity, and inclusion* (DEI) have become central to *corporate social responsibility* (CSR) and *environmental, social, and governance* (ESG) communication. This development has prompted growing

interest in understanding how companies communicate DEI issues to stakeholders, also from a linguistic perspective. In particular, discourse analysts have drawn their attention to how companies construct the notion of DEI and self-represent themselves in relation to the recognition, respect and inclusion of diversity. In this scenario, the inclusion and protection of LGBTQ+ communities – both through employment opportunities and initiatives that support their social well-being locally and globally – have become priorities for internationally competitive companies.

CSR communication is more likely to be perceived as sincere when it is transparent and balances expressions of commitment with substantial descriptions of concrete actions. Such a balance reduces perceptions of opportunism and enhances the credibility of CSR claims. In light of this and given the scarcity of pragmatic and multimodal approaches, this research first aims to examine the linguistic-discursive strategies used to represent LGBTQ+ people and communicate corporate approaches towards their inclusion. In addition, it seeks to identify, from a more pragmatic perspective, the linguistic choices that either facilitate or hinder cooperative dialogue with stakeholders in terms of the transparency of the information disclosed.

Following this introduction, Section 2 reviews research on CSR communication and DEI, from both management (Section 2.1) and linguistic (Section 2.2) perspectives, and introduces pragmatic and multimodal contributions relevant for the assessment of sincere versus opportunistic CSR/ESG communication (Section 2.3). Section 3 illustrates the research aims (Section 3.1), the dataset, and the methodology adopted to analyse the corpus data (Section 3.2). Section 4 presents the main findings, and Section 5 discusses these results, making concluding remarks.

2. Background

2.1. CSR: concept development, LGBTQ+ inclusion, and sincere vs. opportunistic communication

The concept of CSR, introduced in the second half of the previous century, has progressively become a popular object of interdisciplinary research in human and social sciences, especially within management studies and English applied linguistics. Although there is no universally shared definition of CSR, it is commonly referred to as the incorporation of environmental, social, and economic concerns into a company's strategies and operations (e.g., Carroll 1979; Dahlsrud 2008). In more recent years, companies have progressively adopted the notion of ESG, which represents an evolution and specialisation of the concept of CSR: compared to the concept of CSR, which is more philosophical, ESG refers to an approach linked with the quantification and the

measurement of corporate sustainability practices. Despite this difference, which is mainly technical and terminological, both CSR and ESG represent a managerial approach through which companies attempt to address social and environmental challenges while conducting their business. This type of information is included in various genres of corporate communication, such as CSR reports, press releases, and websites, through which corporates show their responsible behaviours to their stakeholders – including investors, employees, and customers (Breeze 2012). In particular, companies have continued to communicate about policies aimed at reducing their negative environmental impact and ensuring social sustainability. In relation to the last aspect, practices for DEI have also been covered to face social changes due to the rise of social movements such as MeToo, Black Lives Matter, disability rights, and LGBTQ+ activism, which is the focus of this paper.

As regards this latter identity group, recent management research has shown growing interest in the ways companies communicate and promote inclusive practices addressed to stakeholders from the LGBTQ+ communities. For instance, Kim (2023) listed three CSR approaches to LGBTQ+ diversity involvement, namely *proactive*, *passive* and *refusal*: the first entails voluntary initiatives beyond legal requirements, demonstrating the company's social commitment to the cause; the second is the adherence to legal requirements without giving proof of an active engagement; the third implies not only a lack of engagement with LGBTQ+ communities, but also support for movements questioning the promotion of this diversity. The scholar further argues that the adoption of specific corporate policies towards LGBTQ+ communities affects stakeholders' perceptions of the company. Kim also maintains that, in general, in geographic and social contexts that are more or less attentive to LGBTQ+ rights, stakeholders tend to have a more positive view of a company when the inclusion of these communities is communicated and perceived as one of its core values and strategic objectives, leading also to stronger purchase intentions.

With respect to the communicative aspect, a sincere and genuine communication of support to LGBTQ+ communities is preferable to a hypocritical and opportunistic approach. In relation to this, Fassin and Buelens (2011, pp. 594-595) identified several types of corporate approaches that can be positioned along a hypocrisy-sincerity continuum, considering factors such as the motives underlying a company's approach, the actual actions implemented in line with it, and the amount of information disclosed to stakeholders. Regarding the communicative aspect, which is the focus of this study, they distinguished between *sincere* and *opportunistic* approaches. The first is characterised by a strong motivation toward societal or ethical values, combined with a high level of actual implementation of actions consistent with those values, and implying a minimal gap between actions and communication. The second refers to cases in which organisations “only demonstrate their ethical

credentials, because this kind of behaviour is perceived, by important stakeholders (e.g. clients), as positive” (p. 595). In this respect, organisations may make false promises, signalling their intention to act in a certain way without necessarily succeeding in or intending to keep them. Moreover, opportunists “communicate intensively but very selectively” (p. 595), being cautious about disclosing information on issues where they are “vulnerable to criticism” (p. 595).

2.2. Linguistic approaches to DEI in CSR and ESG communication: discourses and transparency vs. vagueness

Over the past two decades, discourse analysts have conducted numerous studies on CSR communication, analysing textual and genre features (e.g., Breeze 2012; Catenaccio 2008; Garzone 2005; Malavasi 2017a; Yu and Bondi 2017) and the linguistic-discursive strategies used to construct corporate identity (e.g., Malavasi 2010, 2011), build stakeholder trust (e.g., Fuoli 2018; Malavasi 2017b), and communicate environmental (e.g., Lischinsky 2015) and social sustainability. As for the communication of social issues, particular attention has also been dedicated very recently to DEI (Cifalinò and Catuli 2023; Diani and Turnbull 2025b; Falcone 2025; Iori 2025; Malavasi 2023, 2025; Nocella 2023, 2025; Turnbull 2023; Zaupa 2025).

Findings from this body of research highlight companies’ use of formulaic constructions to promote DEI and to position themselves as reliable, socially responsible actors. More specifically, they reproduce lexico-phraseological patterns through which DEI is constructed as both a corporate value and a strategic goal, outlining intended courses of action to foster these values within the organisation and in society at large. Especially in the disclosures of American-based companies that operate internationally, a holistic approach emerges toward ensuring equal opportunities for employees and other individuals, irrespective of their visible and invisible differences. Zaupa (2025) observed that companies’ linguistic and discursive choices contribute to framing DEI as an integral component of their corporate management. This discursive self-representation is realised through a balanced emphasis between acknowledging the relevance of DEI both within and beyond the organisation, and the disclosure of voluntary actions and policies aimed at ensuring DEI, including targeted recruitment and training initiatives.

While these studies focused primarily on non-financial reports, Cifalinò and Cutuli (2025) extended their analysis to the social media platform X. Their research showed a more dynamic and interactive narrative, engaging audiences through storytelling, advocacy, and awareness campaigns. This suggests that, while reports may serve to reinforce corporate commitments, social media adopt a more immediate, participatory, and engaging communicative approach. The frequent use of narration – more particularly, storytelling – was also observed in

Iori's (2025) study on Instagram posts by "Transport for London" (TfL). She argued that this strategy may be strategic to "encourage both recognition of marginalised voices and collective responsibility for fostering an equitable [...] environment" (p. 115). Regarding the representation of specific diverse identities in the posts examined from a linguistic perspective, to effectively promote an inclusive communicative approach, TfL foregrounded members of specific diversity groups – represented either collectively or individually – while backgrounding other passengers implicitly attributed responsibility for their challenges.

Maintaining a positive reputation with stakeholders requires transparent communication of corporate commitments and practices on social issues (Fombrun and Van Riel 2004; Kim and Lee 2018). The concept of *transparency* is commonly understood as the open sharing of information, but it has also been conceptualised as information "amount" (Winkler 2000) or "completeness" (Vaccaro and Madsen 2006). Schnackenberg and Tomlinson (2016) proposed a model of transparency comprising three dimensions, namely *disclosure*, *accuracy*, and *clarity*. The first dimension concerns the completeness, relevance, and timely communication of information; the second refers to its correctness and reliability; the third to its comprehensibility for the receiver.

However, transparency is also subjective, as its degree depends on the perceptions of both the writer and the receiver, making it difficult to identify universally shared linguistic markers of transparency. In a similar way, linguistic vagueness can be either intentional (Channel 1994) or interpretable depending on the context in which it is reproduced (Cheng and Warren 2003). In some cases, vague language is also strategic to adjust the level of detail according to the communicative context (Koester 2007). Jin (2022) aligned with this idea on the basis of a corpus-based analysis of CSR reports from companies in the cosmetics sectors, providing a taxonomy through which vague language markers may be identified. She argued that these choices may offer communicative flexibility in contexts of uncertainty or potential criticism (p. 77).

In the past few years, an increasing number of corpus-based studies have explored transparency and vagueness in CSR and ESG disclosures (e.g., Bondi and Sezzi 2025; Crawford Camiciottoli 2025; Diani and Turnbull 2025a; Malavasi and Nocella 2025; Poppi 2025; Zaupa 2024). In the context of DEI communication, this aspect was a complementary research object in most of the studies reviewed in the first part of this section. The findings of these studies showed a common corporate approach to presenting DEI initiatives yet including limited information on their actual implementation (Diani 2025; Nocella 2025; Zaupa 2025), thereby reducing the effectiveness of transparent communication (Iori 2025, p. 115) or even suggesting a lack of it (Zaupa 2025). While research has shown that companies tend to adopt a balanced communicative approach focused on information completeness and clarity

(Malavasi 2025), the frequent use of vague linguistic expressions challenges this apparent degree of transparency (Zaupá 2025).

The studies discussed in this section have shown an increasing interest in the linguistic choices through which companies communicate their approach to DEI, whether transparently or vaguely. However, these contributions have not focused on each specific diversity aspect. Moreover, further research is needed to shed light on the linguistic and discursive strategies that may suggest either a sincere or an opportunistic approach to the inclusion and promotion of diversity. For this purpose, the following section seeks to integrate this linguistic and communicative background with insights from pragmatics, providing additional tools for the study of CSR and ESG communication.

2.3. Pragmatics and multimodality for sincere vs. opportunistic social in/ex-clusion in CSR and ESG disclosures

Although CSR communication does not occur in a conversational setting due to its asynchronous nature, it can nevertheless be considered as a communicative exchange between companies and their stakeholders, given that one of its primary purposes is to maintain trust relationships. This conceptualisation implies that companies should select appropriate linguistic resources and strategies to engage in a form of cooperation with their addressees. This aligns with what Grice (1989) referred to as the *Cooperative Principle*. According to Grice, such a cooperation is ensured through the respect of conversational maxims – *quantity*, *quality*, *relation*, and *manner* – which may also be associated with Schnackenberg and Tomlinson's (2016) dimensions of transparency. According to the maxim of quantity, speakers should provide as much information as required, which can be associated with the transparency dimension of disclosure in terms of information amount. The maxim of quality refers to the truthfulness of utterances, linking with the accuracy and reliability of information. The maxim of relation may be linked to the transparency dimension of disclosure as well, in terms of relevance of information. Ultimately, the maxim of manner involves avoiding obscurity, ambiguity, and disorder, which may correspond to the clarity of information.

The identification of communicative strategies in CSR disclosures and the assessment of the level of transparency of the information conveyed through them may offer useful tools for determining whether a company's communication of its commitment to a specific issue tends to be opportunistic or sincere. Drawing on the studies discussed above, the former is likely to rely primarily on commissive statements and linguistic choices that foreground corporate identity and vision rather than concretely implemented measures. The latter, by contrast, would be characterised by a more fact-based approach. However, reporting actions does not necessarily imply their actual implementation. The degree of transparency of information about such actions

may therefore serve as an indirect indicator of the alignment between declared commitments and reported practices, helping to assess whether a minimal gap exists between communication and implementation - an essential condition for identifying sincerity in CSR communication. A lack of transparency, in turn, may call into question the sincerity of the disclosed information, suggesting a more opportunistic communication.

Beyond linguistic-discursive and pragmatic choices, the selection of images can also contribute to specific discursive representations and establish a visual relationship with the addressee, shaping perceptions of the degree of genuineness or artificiality of the communication (e.g., Catellani 2015). In the context of CSR communication on DEI, it has been argued that the randomized integrations of images in CSR disclosures may reinforce specific diversity narratives, but also existing social stereotypes (e.g., Singh and Point 2006), as well as support linguistically vague information. Zaupe (2025) attempted at examining the interplay between text and integrated images through a visual analysis of a small sample of images integrated in CSR reports and corporate websites, using an ad-hoc analytical framework that revisits and adapts categories from van Leeuwen's (2008) theory of the visual representation of social actors. His study shows a balanced representation of employees from different ethnic backgrounds, reflecting the influence of recent social movements on corporate communication. According to him, the frequent use of images shot at close distance may suggest closeness of the companies with their stakeholders, reinforcing the linguistic choices employed to express their commitment to DEI. At the same time, the inclusion of generic subjects and homogeneous groups visually reinforces the holistic and sometimes vague communicative approach to DEI observed at the linguistic level.

3. Research aims, materials and methods

3.1. Research questions

In light of the limited linguistic research examining companies' discursive strategies to address LGBTQ+ people and issues in their CSR and ESG disclosures, along with the absence of studies adopting both pragmatic and multimodal perspectives, this paper seeks to answer the following interrelated research questions (RQs):

- RQ1. Which people from LGBTQ+ community are most visible in CSR and ESG disclosures, and how are they represented discursively?
- RQ2. Which communicative strategies do companies reproduce to disclose information about LGBTQ+ inclusion?
- RQ3. Is this information communicated transparently or vaguely, and which conversational maxims are primarily observed to convey one or the other?

- RQ4. How are people and communities visually represented as social actors in the context of CSR and ESG disclosures about LGBTQ+ inclusion?
- RQ5. To what extent does the analysis of linguistic and visual resources support an assessment of the degree of sincerity or opportunism underlying corporate communication?

3.2. Data and methodology

To address the research aims, this study analyses an exploratory corpus of 77,316 tokens, consisting of English-written CSR and ESG disclosures published between 2020 and 2023 – these include non-financial reports, press releases, and ESG webpages. As for the reports, given the study’s focus on LGBTQ+ support and inclusion, only the sections or paragraphs explicitly addressing this information were isolated and included in the corpus.

Following the criteria of data availability and heterogeneity of sectors, the dataset created includes texts published by 11 American-based companies that operate internationally, each operating in a different sector. At the time of the selection, they ranked first among U.S. employers within their respective sector in *Forbes*’ 2023 list of “Best Employers” for their respective sectors.¹ The choice to rely on data from this magazine is justified by the fact that this ranking, developed in cooperation with the global data and business intelligence platform *Statista*,² is based on large-scale surveys where employees rate their employers on factors such as work environment, salary, conditions, career development, diversity, and reputation.³ It was further checked whether the selected companies were included in the *Human Rights Campaign*’s “Equality Index”, a ranking that annually identifies the best employers in the U.S. according to their commitment to implementing LGBTQ+ inclusive policies and practices.⁴ Below (Table 1) a breakdown of the corpus is provided.

¹ Although the 2023 ranking is no longer available online, the companies selected were again listed among *Forbes*’ Best Employers in 2024, albeit in different positions. See: <https://www.forbes.com/lists/worlds-best-employers/> (20.07.2025).

² See: <https://www.statista.com/> (last visited: 20.07.2025).

³ See: <https://r.statista.com/en/employers/worlds-best-employers-2024/ranking/#methodology> (20.07.2025).

⁴ See: <https://reports.hrc.org/corporate-equality-index-2023#rating-system-methodology> (15.08.2025).

Sector	Company	Tokens
Food and beverage	Coca-Cola	2,840
Automotive	Ford	996
IT & Software	Microsoft	5,139
Electronics	IBM	8,095
Hotels	Marriot	4,365
Media	Netflix	1,552
Clothing	Nike	3,647
Banking	PayPal	1,104
Pharmaceutical	Pfizer	25,663
Restaurants	Starbucks	21,455
Retail	The Home Depot	3,084

Table 1

Corpus structure: sectors, companies and tokens.

Once compiled, the corpus was uploaded and processed on the software Sketch Engine (Kilgariff *et al.* 2014) for the linguistic analysis. The dataset was first linguistically investigated by adopting a synchronic corpus-based and driven approach (Tognini-Bonelli 2001), and this level of analysis involved two phases, quantitative and qualitative.

The quantitative phase consisted in generating the frequency list through the function *Wordlist*, with the aim of shedding light on the mostly represented people from the LGBTQ+ community. Given the focus of this research on gender and sexual identities, only words referring to these identities were subsequently selected from the 100 most frequent words, and their lexical collocates were calculated through the *Word Sketch* tool. This function generates the collocates of any search word (including its plural form) and displays them together with the score signalling their collocational strength with the search word. Moreover, it also shows the distribution of the collocates across different grammatical categories, such as the modifiers of the word, the nouns it modifies, and the verbs where it appears as subject or object. Given that significant discourses can be unveiled from repeated and not isolated lexical patterns around specific objects, people, or concepts (Baker 2006), this analysis considered the lexical collocates occurring with the search words at least 3 times.

The qualitative phase involved the analysis of collocates in context through close reading of their extended concordance lines. These lines were obtained by aggregating all concordance lines generated by the software for each collocate, excluding duplicate concordance lines of collocates co-occurring in the same text. This analytical stage was aimed at exploring the most salient lexico-phraseological patterns (Sinclair 2004) around these collocates, to reveal the companies' most salient strategies to communicate their approach to the inclusion of LGBTQ+ communities. These were identified drawing on Yu and

Bondi's (2017) rhetorical strategies in CSR reports. While originally conceptualized as genre-structuring moves, in this study they are readapted as strategies through which companies communicate their commitment to specific CSR issues. The table below (Table 2) illustrates the strategies that were investigated during the close reading of the extended concordance lines.

Macro-strategy	Strategy	Sub-strategy
Presenting the company	Presenting corporate profile	
	Presenting corporate governance	
	Stating values and beliefs	
	Stating missions	
	Showing commitment	
	Establishing credentials	
Presenting the company's CSR strategy	Stating strategies / methods / practices	
	Previewing future performance	
Reporting on CSR activities and performance	Presenting performance	Reporting actions and results
		Communicating strategies/methods/practices
		Stating missions
		Assessing performance
		Detailing an internal action
	Presenting an internal action	
	Detailing an issue	
	Presenting individual cases	Action
Situating the context	Describing external circumstances	Achievement
		Experience
		Referring to external circumstances
	Introducing an aspect of CSR performance	Stating general features and universal laws
		Demonstrating importance of an issue
	Presenting risks and difficulties	

Table 2
Communicative strategies readapted from Yu and Bondi (2017).

To pursue the complementary aim of assessing the degree of transparency of this information and, thus, its level of sincerity or opportunism, the close reading of the concordance lines also involved the identification of the lexico-communicative choices suggesting compliance with Grice's (1991) maxims, or vagueness of information. For the latter, the analysis focused on the identification of Jin's (2022) vague expressions related to quantity (e.g., *more than, many*), time and frequency (e.g., *recent, often*), and degree (e.g., *important, significant*). Other vague language markers that were detected include Channel's (1994) generic nouns (e.g., *things, activities*) and vague approximators (e.g., *about, around*), as well as quality-assessment expressions (e.g., Zaupe 2024), namely evaluative adjectives, verbs, or adverbs that are not preceded or followed by contextual information justifying the companies' evaluation.

Given the complementary aim of this paper to examine whether the choice of visuals support or contrast with the findings from the verbal analysis, a total of 31 images depicting human subjects and embedded in the disclosures included in the dataset were manually extracted for the visual analysis. The analysis drew on an adapted and reduced version of van Leeuwen's (2008) framework for the visual representation of social actors. Specifically, a selection of categories from his model was retained in order to identify visual patterns that were effectively observable, given that some images were embedded in the disclosures without captions or accompanying information that would allow them to be linked to specific portions of text. As suggested by the scheme below (Figure 1), two levels of analysis were conducted, here referred to as *identification and visualisation*, and *visual relationship with the viewer*.

The first level of analysis attempts identify the represented social actors and the ways in which they are presented to the viewer. At this stage, two of van Leeuwen's dimensions were considered: *specific* vs. *generic* social actors, and *individual* vs. *group* representation. The former concerns whether individuals are represented with identifiable features (such as physical traits that signal a particular ethnic background), enabling viewers to associate them with a defined social group, or instead portrayed as generic and decontextualized subjects whose identities remain unspecified. Concerning the latter dimension, social actors may appear either as single individuals or as members of a group. In collective representations, two further sub-strategies can be observed. One is *differentiation*, whereby individuals are shown interacting with others while still displaying distinctive characteristics that signal their belonging to a particular identity group. The other is *homogenization*, in which multiple individuals, despite showing visible differences (for example, in terms of geographic origin), are portrayed as a collective actor, thereby obscuring or backgrounding their individual identity traits.

The second level of analysis aims to examine the visual relationships established between the represented subjects and the viewer, relying on the

dimension of *distance* and *relation*. The former concerns the degree of spatial proximity between the portrayed actors and the viewer, which contributes to shaping the type of relationship constructed: *close* shots tend to evoke intimacy and engagement, whereas *far* shots convey distance and detachment. The latter dimension is further subdivided into three sub-categories: *involvement*, *power*, and *interaction*. Involvement and power are primarily communicated through the angle from which the image is taken. A frontal angle typically conveys a sense of *involvement*, while an oblique angle may signal *detachment*. As for *power*, a high angle positions the viewer as dominant over the represented actor, whereas a low angle gives greater prominence to the latter; an eye-level perspective, instead, suggests a relationship of equality. Finally, *interaction* is expressed through gaze direction. When social actors look directly at the viewer (*direct address*), a form of visual engagement is established. By contrast, an averted gaze indicates *indirect address* and tends to create a sense of distance or reduced involvement.

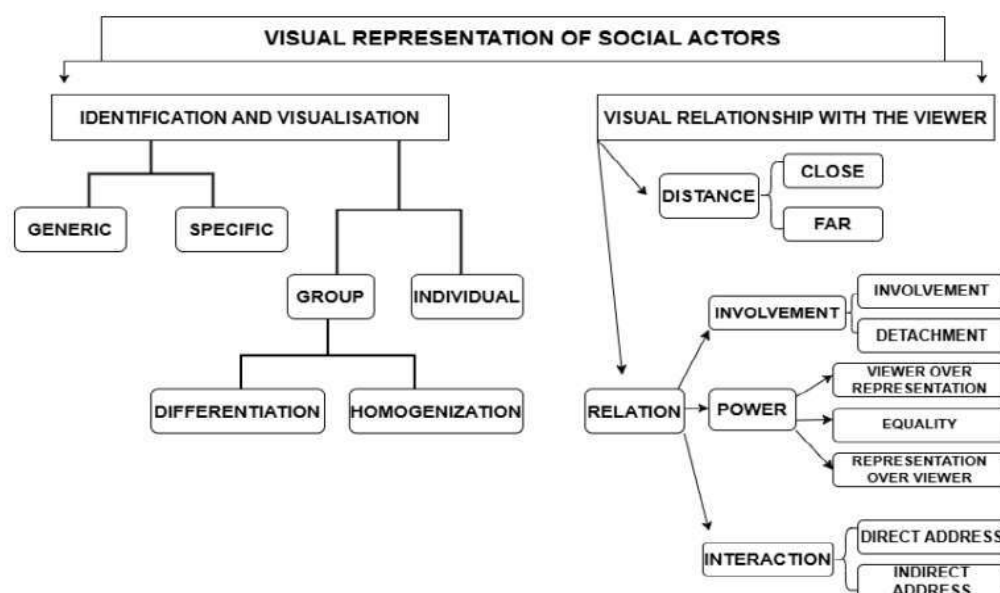


Figure 1

Visual analysis readapted from van Leeuwen (2008, in Zaupa 2025, p. 129).

For the sake of clarity, in both the linguistic-pragmatic and visual analyses, multiple manual annotations were assigned to each of the analytical categories outlined above. Moreover, instances showing more than one discourse strategy, maxim, or marker of vague language were assigned multiple tags. Therefore, the quantitative results presented in the following section should be interpreted considering that frequencies refer to each individual strategy, linguistic or visual pattern, out of a total of 105 linguistic instances or 31 images.

4. Results

4.1. Identities' representations

Starting from the analysis of the most frequent words, as shown in the table below (Table 3), it was observed that among the 100 most frequent words denoting LGBTQ+ identities are the acronym *LGBTQ* itself, and the words *gender* and *transgender*.

Word	Frequency	Normalised frequency (per 1,000 words)
LGBTQ	288	3,72
GENDER	93	1,20
TRANSGENDER	77	1,00

Table 3
Most frequent nouns related to LGBTQ+ identities.

Among these three terms, the acronym LGBTQ occurs approximately three times more frequently than *gender* or *transgender*. This frequency pattern may reflect a corporate policy that adopts a more holistic, rather than targeted, approach to gender and sexual diversity, as further evidenced by both the quantitative and qualitative analysis of the collocates discussed in the following paragraphs.

Table 4 below illustrates the lexical collocates of these three terms, from which a total of 105 concordance lines were extracted and subsequently read for the qualitative analysis conducted in this study.

Word	Nouns and adjectives	Lexical verbs
LGBTQ	WOMAN (11,5); VETERAN (10,9); EQUALITY (10,8); DISABLED (10,6); COMMUNITY (10,5); PEOPLE (10,4); LOCAL (10,1); PARTNER (9,3)	SUPPORT (11,5); FIND (10,6); CELEBRATE (9,9)
GENDER	IDENTITY (12,8); RACE (12,1); SURGERY (10,8); EQUITY (10,3); EXPRESSION (10,3)	None
TRANSGENDER	BISEXUAL (12,0); GAY (11,9); PARTNER (11,4); EMPLOYEE (10,5); PEOPLE (9,7)	None

Table 4
Lexical collocates of *LGBTQ*, *gender* and *transgender*.

The generic and collective corporate approach towards LGBTQ+ identities discussed above may also be inferred from the most frequent co-occurrence of both generic words referring to social actors – such as *community*, *people*, and

partner – and more specific identity groups, namely women (*woman*), the elder (*veteran*), disabled people (*disabled*), local communities (*local*), ethnic groups (*race*), and other sexual orientations (*bisexual* and *gay*).

The analysis of the collocates in their linguistic context further confirms this discursive collectivisation of LGBTQ+ people as social actors. In 19 of the 105 concordance lines, the collective co-occurrence of the collocates referring to different identities was observed within instances reporting corporate practices (see Section 4.2) to ensure social inclusion of LGBTQ+ people, as shown in the excerpts below: (1) refers to The *Home Depot*'s support to businesses owned by people belonging to other diversity groups (*women*, *veteran*, and people with *disability*); (2) points to employment opportunities, still mentioning a wide range of social actors, as suggested by the co-occurrence of the collocates *race*, *gender identity*, and *veteran*.

(1) The Home Depot's Supplier Diversity strategy includes ethnic-minority, *women*, **LGBTQ+**, *veteran* and *disability*-owned business partners. (The Home Depot - 2023 ESG Report)

(2) All qualified applicants/employees will receive consideration for employment without regard to the individual's age, *race*, color, religion or creed, national origin or ancestry, sex (including pregnancy), sexual orientation, **gender**, *gender identity*, physical or mental disability, *veteran* status, genetic information, ethnicity, citizenship, or any other characteristic protected by law. (Coca-Cola webpage: Equal Employment Opportunity Employer)⁵

Regarding the most represented stakeholders from the LGBTQ+ community, the strong co-occurrence of *community* with *LGBTQ* and *employee* with *transgender* suggests a corporate commitment to social engagement with local LGBTQ+ communities, as well as an approach aimed at fostering the social and economic well-being of trans employees. This dual focus in corporate policies also emerged from the analysis of the concordance lines. However, a greater number of references to the support of LGBTQ+ and trans employees (63 instances) than with external communities (34 instances) was observed.

The verbal collocates of *LGBTQ+*, i.e., *support* and *celebrate*, may point to an ideologically shared corporate policy that positively impacts the lives of LGBTQ+ communities. The discursive construction of LGBTQ+ people as beneficiaries of the various inclusive practices and initiatives could be further supported by 32 concordance lines showing the frequent co-occurrence of lexical items belonging to the semantic field of *care* – these include, for instance, the use of verbs such as *defend* and *protect*, as shown in the examples below.

⁵ Source: <https://www.coca-colacompany.com/policies-and-practices/equal-employment-opportunity> (31.12.2025).

(3) Global Equality Fund Marriott continued its support for the Global Equality Fund (GEF), a public-private partnership dedicated to advancing and *defending* human rights and the fundamental freedoms of **LGBTQ people** globally. (Marriot – 2021 ESG Report)

(4) Starbucks also joined the Human Rights Campaign's Business Coalition in support of the Equality Act, a bill that would *protect* **LGBT people** from discrimination in housing, the workplace, public accommodations, and other settings under federal law. (Starbucks – Press Release: Equity, Inclusion and Diversity Timeline. 01.08.2022)

4.2. Communicative strategies

Moving on to the communicative and discursive strategies used by the companies to address LGBTQ+ stakeholders, the table below (Table 5) shows the distribution of the communicative strategies identified through the reading of the extended concordance lines.

Macro-strategy	N	Strategy	N	Sub-strategy	N
Presenting the company	20	Stating values and beliefs	4		
		Stating missions	1		
		Showing commitment	15		
Presenting the company's CSR strategy	11	Stating strategies / methods / practices	9		
		Previewing future performance	2		
Reporting CSR activities and performance	84	Presenting performance	63	Reporting actions	48
				Reporting results	12
				Stating missions	2
				Assessing performance	2
		Presenting individual cases	20	Experience	20
Situating the context	22	Describing external circumstances	10	Demonstrating importance of an issue	10
		Introducing an aspect of CSR performance	12		

Table 5
Communicative strategies.

Findings first suggest that companies tend to address stakeholders primarily through a factual communicative approach, as reflected in the high frequency of segments tagged with the macro strategy *Reporting CSR activities and performance* in relation to LGBTQ+ inclusion (84 instances). However, the analysis simultaneously reveals a substantial imbalance between references to corporate initiatives and evidence of tangible outcomes, with significantly more emphasis placed on *actions* (48 instances) than the *results* (12 instances) achieved to demonstrate their effectiveness.

With regard to the analysis of the segments tagged under the category *reporting actions*, two differences can be observed in the linguistic contexts of *LGBTQ* and *transgender*. On the one hand, various references to corporate practices can be found, including the allocation of financial resources and participation in legal initiatives aimed at promoting the inclusion of external LGBTQ+ communities or organisations, as well as employees (see examples (5) and (6)). On the other hand, most of the instances of *transgender* primarily refer to corporate practices related to financial health benefits provided to transgender employees (see example (7)).

(5) The Starbucks Foundation announces it has awarded 100 grants to nonprofit **LGBTQ+ communities**, totalling more than \$300,000. (Starbucks – Press Release: Timeline: Starbucks history of LGBTQIA2+ inclusion, 16.05.2022)

(6) We have *endorsed* the *Equality Act* to ensure fair treatment and workplace protections under the law for members of the **LGBTQ community**. (Marriot – 2021 ESG Report)

(7) At the heart of this work is ensuring we have inclusive benefits – like family expansion reimbursement for adoption and surrogacy, expanded *health insurance* options for **transgender partners**, and *medical travel reimbursement* when partners are legally unable to access *gender affirming care*. (Starbucks – Press Release: A Message from Starbucks, 23.06.2023)

As for the reading of the concordance lines tagged with the strategy *reporting results*, it was observed that this strategy, besides serving as evidence of the implementation of corporate practices, is strategically used as a self-promotional tool to emphasise its positive outcomes. Example (8), referring to *Starbucks'* recognition as an employer ensuring LGBTQ+ equality, is a case in point.

(8) In FY21, Starbucks *received a 100% score* on the Human Rights Campaign Corporate Equality Index based on corporate policies and practices for **LGBTQ equality** for the 11th year. (Starbucks – 2021 Global Impact Report)

The example above also shows that the reported results are accompanied by brief contextual information concerning the recognition received by the company – in this case, a description of the *Human Rights Campaign Corporate Equality Index*. As this pattern occurs in all the concordance lines with the tag

Reporting results, the frequency of this strategy coincides with that of *Introducing an aspect of CSR performance*.

Another significant finding emerging from the analysis of the concordance lines with the tag *Reporting CSR activities and performance*, is the occurrence of segments classified as *Presenting individual cases*. Through this strategy, corporate initiatives are narrated from the perspective of the individuals involved. Example (9) below introduces a coming-out experience of an employee of IBM, with the aim of presenting the company as an inclusive environment.

(9) Initially, I had no intention of ever coming out in a professional setting. My plan was to explore my **gender identity** in my personal life and continue to identify as a man in my professional life. As I continued exploring my gender identity, it became clear that the best thing to do was to be completely honest with everyone in my life. I decided to come out as transgender to everyone I have contact with in both my personal and professional lives. (...) It is a moment in my IBM career that I will remember forever. Our team has a reputation for being a “family”, and I was proud if our culture played even the slightest role in Claire feeling comfortable enough to live her professional life as her authentic self. (IBM webpage: Gender transition)⁶

Although the macro-strategies *Presenting the company* and *Presenting the company's CSR strategy* are less prevalent than the factual strategies discussed above, the data likewise reveal that companies communicate LGBTQ+ inclusion framing this issue not only as a key driver of their CSR strategy but also as a core organisational value. As shown in the table above, the discursive strategy most frequently adopted in this case is the commissive statement, through which promises or commitments are made to act by implementing tools and policies aimed at ensuring inclusive and equal opportunities for LGBTQ+ people. More particularly, the analysis of the segments tagged with *Showing commitment* reveals that this strategy is reproduced through the explicit use of the verb *commit* (see example (10)), verbs expressing continuity, such as *continue* (see example (11)), as well as lexical choices conveying effort and necessity, such as the verbs *strive* in (10) and *urge* in (12).

(10) U.S. PARTNER NETWORKS Strive to cultivate an equitable, dynamic and supportive environment for **LGBTQ partners**, allies and customers. Global Human Rights at Starbucks, *we are committed* to respecting human rights, as outlined in our Global Human Rights Statement. (Starbucks – 2022 Global Impact Report)

(11) Since launching our gender transition support framework and corresponding white paper in 2017, IBM *has continued* to evolve its support of **transgender employees**. (IBM – 2021 Impact Report)

⁶ Source no longer available on the web.

(12) June 2016 Starbucks joins more than 200 other business leaders to *urge* equal treatment for the **LGBTQ community**. (Starbucks – Press Release: Equity, Inclusion and Diversity Timeline, 01.08.2022)

Finally, noteworthy are the ten instances classified as examples of the macro-discourse strategy *situating the context*, in which companies refer to the broader social context in which LGBTQ+ communities live – particularly the challenges they face – in order to reiterate the importance of issues related to their inclusion. Example (13) is a case in point.

(13) Too many LGBT+ people can't show up as their authentic selves at work without facing negative consequences,” said Deena Fidas, Chief Program and Partnerships Officer for Out & Equal. “Too many can't express their true gender identity”. (IBM – Press Release, 09.07.2021)

The presence of this strategy suggests both an attempt of the companies to reinforce their perceived genuineness of inclusive practices and a communicative choice aimed at responding to the pressures caused by the social movements mentioned above.

4.3. Transparency vs. vagueness

The following table (Table 6) presents the results that emerged from the final level of the linguistic analysis - the transparent or vague communication of companies' practices and results - specifying the number of instances displaying with each maxim, as well as the markers of vagueness identified during the analysis.

Maxim	No. of instances	Marker of vagueness	No. of instances
Quantity	47	Nouns with generic meaning	16
Quality	18	Vague quantity expressions	21
Relation	33	Quality assessment expressions	4
Manner	14		

Table 6
Transparent vs. vague communicative choices.

Data suggest that in more than one-third of the 105 extended concordance lines the companies' linguistic choices seem to enhance the perception of information as complete, accurate, reliable, and relevant. This outcome suggests a simultaneous respect of the maxims of quantity, quality, and relation, and, consequently, of the transparency dimensions of disclosure and accuracy. In the following extract (see example (14)), for instance, companies use numbers to provide precise information on the results achieved in terms of financial resources donated to LGBTQ+ organisations.

(14) *2021* Supporting vulnerable members of the **LGBTQ+ community** Starbucks donates \$50,000 to the Lavender Rights Project to support providing low-cost *legal services* for transgender and queer low-income people and other marginalized communities in Washington state and \$ 50,000 to the National Center for Transgender Equality to increase understanding and acceptance of trans people across the United States. (Starbucks – 2021 Global Impact Report)

(15) In addition, each of our eight Employee Networks donates \$25,000 annually to *nonprofit organizations aligned with their goals*. Intersectionality across **race, gender**, and orientation is a key consideration in our grant making, as it enables us to create more impact in each of the communities we support – individuals of all abilities; Asian American, Middle Eastern and Pacific Islander communities; the Black community; the Latinx community; our military and veterans communities; Native American and Indigenous communities; the **LGBTQIA+ community**; and girls and women. (Nike – 2020 Impact Report)

Beyond the use of numerical data, which conveys completeness of information, references to project names, beneficiary organisations and identities (see examples (14) and (15)), as well as specific years (e.g., *2021* in (14)), enhance the relevance and reliability of information, adhering to the maxims of relation and quality. More specifically, information in (14) is disclosed in a timely manner because it refers to a practice of 2021 within disclosures of the same period; in (15), the specification of the beneficiaries enhances information relevance, as such details are meaningful for specific communities who could read the disclosures to evaluate the company's inclusiveness towards them. Regarding the latter, the precise quantification of financial resources invested, and the specific names of initiatives and beneficiary organisations suggest that this type of information is highly likely to be accurate, as it would otherwise be subject to criticisms by the social actors involved.

Finally, although they were very scarce, few instances were also identified among the concordance lines, where information was disclosed through linguistic choices conveying clarity, in line with the maxim of manner. These mostly involve the use of exemplification, as suggested by the linker *such as* in the example below.

(16) Supporting local **LGBTQ+ communities** in Brazil Starting in 2018, Starbucks in Brazil has worked with local advocacy groups, *such as* Casa 1 (Sao Paulo) and Grupo Arco-Iris (Rio de Janeiro), among others, to **support LGBTQ+** inclusive inclusion efforts through the donation of Starbucks product sales *such as* the Pride Frappuccino® and t-shirts, to help those in vulnerable situations due to their sexual orientation. (Starbucks – Press Release: Timeline: Starbucks history of LGBTQIA2+ inclusion, 16.05.2022)

However, in some instances, the lexical items conveying transparency also co-occurred with choices that challenged the degree of information transparency: in (14), for instance, *legal services* is used generically in its context, without describing the services offered; similarly, in (15), the noun *organization* in the expression *to nonprofit organizations aligned with their goals* is vague since the names of the organisations benefitting from the company's donations were not made explicit in the preceding or following textual context. Moreover, in numerical reporting in the example below, some numbers are precise (e.g., 155, 122, 28, 79), while others are approximate (*over 900*) or expressed with high percentages that foreground success and obscure deficiencies or failures. For example, the use of 99% highlights the positive outcomes of *Marriot* while backgrounding the remaining 1%, which is left unexplained.

(17) In 2020, Marriott's support helped IE screen, vet, and directly represent or place 155 clients with pro bono counsel, enabled staff to answer over 7,000 calls for help to their international hotline, supported attorneys in answering *over 900* international inquiries from **LGBTQ people** fleeing persecution, and sponsored in-depth intake interviews with 122 potential clients. Marriott's support helped enable IE to serve their community at an especially crucial time, as the COVID-19 pandemic caused IE's volume of hotline calls to double. Despite this, IE was still able to support LGBTQ individuals from 28 states and 79 countries, maintaining a *99%-win rate* in their cases. (Marriot – 2021 ESG Report)

Such examples may suggest that, on the one hand, companies seemed to establish communicative 'cooperation' (Grice 1991) with their stakeholders, as shown by adherence to the maxims of quantity, quality, and relation. On the other hand, this is reduced when transparency could expose them to criticism. Other instances of vague language included examples of quality-assessment lexical items, as exemplified by the use of *bigger impact* in the excerpt below.

(18) While we currently track and measure progress against gender globally and race and ethnicity in the United States, our inclusion efforts worldwide span five dimensions: **gender identity**, culture and heritage, generation and life experience, LGBTQ+, and ability and wellness We believe we can make a *bigger impact* toward a more inclusive world through collaboration. We recently joined The Valuable 500, a global movement to advance disability inclusion, and we signed on to the Partnership for Global LGBTI Equality, a coalition of organizations committed to accelerating LGBTI inclusion in the workplace and in communities. The partnership is supported by the Office of the United Nations High Commissioner for Human Rights and operates in collaboration with the World Economic Forum. (Coca-Cola - 2020 Business and ESG Report)

In this example, *Coca-Cola* explicitly states that it has made progress towards inclusion. However, by using the expression *bigger impact*, it offers only an evaluative claim supported merely by references to partnerships with movements engaging to the issue. The lack of descriptions about the practices

and initiatives undertaken within these partnerships undermines the reliability of the claim and provides the addressee with no basis to assess the effectiveness of the company's approach to inclusion.

4.4. Visual patterns

Moving on to the visual analysis of the integrated images, regarding the *identification and visualisation* level, the graph below (Figure 2) shows that representations of LGBTQ+ social actors as individuals, or through differentiation (for example, displaying couples of presumably homosexual partners), are more frequent than those of homogeneous groups.

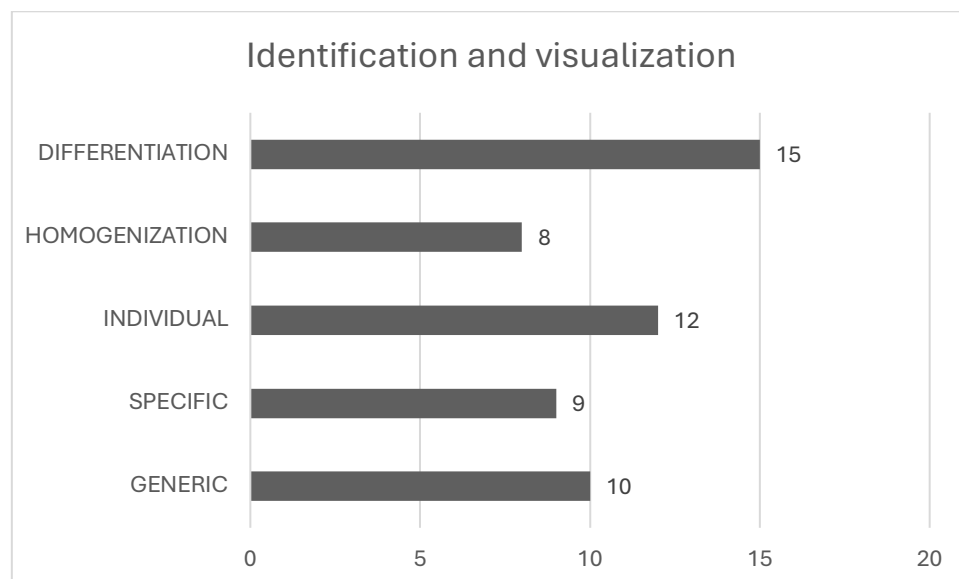


Figure 2
Results from the visual analysis – First level of analysis.

These visual choices may serve as a tool to reinforce diversity narratives. For instance, in the figures below, the explicit portrayal of openly transgender individuals - as inferred by the caption (Figure 3) - and a couple of women presumably in a domestic setting (Figure 4) could be interpreted as a corporate strategic attempt to establish a connection with stakeholders belonging to these specific homosexual and transgender communities.



Figure 3
Individual representation of an openly transgender person.



Figure 4
Couple representation of women.

By contrast, Figure 5 depicts different individuals participating in LGBTQ+ initiatives - as suggested by the rainbow flag displayed above them - as part of a visually homogeneous group. While this representation may signal inclusiveness, it may also make it more difficult for members of specific LGBTQ+ communities to identify with the company's goals and policies, as no clear visual markers distinguish particular gender or sexual identities.



Figure 5
Group representation of people at a LGBTQ+ inclusive event.

In addition, the balanced representation of both specific and generic people further weakens a corporate communicative approach centred on diversity narratives. In a similar way to the previous image, Figure 6 portrays a heterogeneous group of individuals participating in a Pride event. The participants can reasonably be presumed to be company employees, as they are wearing identical T-shirts displaying the same slogan, which is also reproduced on the banner they are holding and which includes the company's name. On the one hand, this visual strategy may reinforce the company's inclusive stance towards LGBTQ+ workers. On the other hand, the absence of additional visual cues linking the individuals included in the image to specific sexual orientations or gender identities limits the visibility of the particular social actors who might be the intended beneficiaries of the company's inclusive policies.



Figure 6
Representation of people participating in a Pride event.

The analysis level on *visual relationship with the viewer*, as shown in the bar chart below (Figure 7), shows that direct address, equality, involvement, and close distance are prevalent in the dataset.

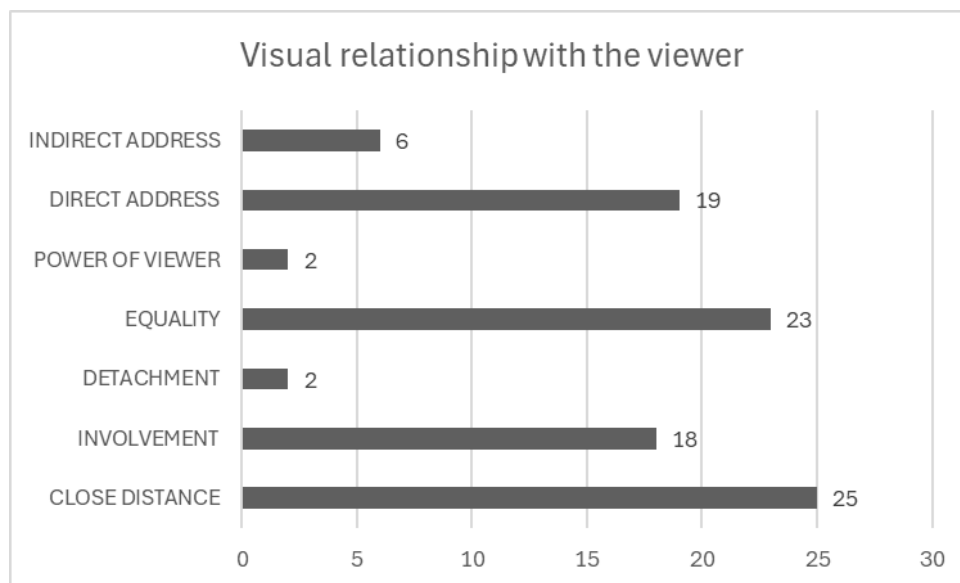


Figure 7

Results from the visual analysis – Second level of analysis.

Most of the images discussed above and the figure below (Figure 8), for example, represent employees at a close distance and looking directly at the viewer, suggesting a relationship of equality between the two interlocutors – the company, represented through its employees, and the potential stakeholders interfacing with the image.



Figure 8

Representation of employees at a Pride event.

While these visual choices may suggest an intention of the companies to engage with LGBTQ+ stakeholders to convince them of the corporate entity's

commitment to inclusion at the workplace, they do not necessarily demonstrate genuine commitment, as they may equally reflect a strategically designed – and potentially artificial – communication.

Finally, another significant visual pattern is the recurrence of symbols associated with LGBTQ+ communities: 13 of the images analysed also feature the rainbow flag, and 3 include the transgender flag. The reproduction of these symbols may also be read as a strategically inclusive corporate policy expressing alignment with LGBTQ+ symbols and, by extension, with the values and political claims they represent.

5. Discussion and concluding remarks

This study analysed, from a linguistic-discursive, pragmatic and multimodal perspective, how inclusion and support for LGBTQ+ people were communicated in a corpus of CSR disclosures from 2020-23 of a sample of US-based international companies.

Two main divergent linguistic-discursive patterns were observed when referring to the LGBTQ+ community and transgender people, respectively. As for the former, companies focus on communicating initiatives for external communities, such as financial support and institutional partnerships, to improve their social well-being. Across the instances analysed, the use of *collectivisation* was observed, as shown by the frequent co-occurrence of terms referring to other minority groups. This finding may suggest a holistic rather than targeted communicative approach for each specific sexual or gender identity, unlike what emerged in the linguistic context of the collocates of the word *transgender*. In this case, the close reading of the extended concordance lines revealed a more concrete focus on workplace equity and health-related benefits for transgender employees. Despite these two different linguistic-discursive patterns, both LGBTQ+ communities and transgender people are mainly represented as beneficiaries of corporate policies, while companies position themselves as ‘proactive’ (Kim 2022) promoters of inclusion.

Another key finding concerns the communicative strategies adopted. Companies tend to highlight implemented actions and achieved results rather than focus on their vision and CSR strategy. Companies also seem to provide complete, accurate, and relevant information, in line with the maxims of quantity, quality, and relation. This is reflected in the frequent use of precise numerical data, explicit references to beneficiary organisations and initiatives, and the temporal alignment between the actions reported and the reporting year of the disclosure, which enhances the relevance of the information provided. However, when the co-text window around each concordance line was expanded, the analysis revealed instances of vague language – including

quantity approximators and lexical items with generic meaning – weakening the degree of information transparency.

From a visual perspective, most of the integrated images analysed portrayed employees, usually at a close distance and with direct gaze, creating an impression of proximity and engagement with the viewer. Regarding the representation and visibility of LGBTQ+ identities, many images depict individuals or couples, sometimes with identifiable gender or sexual identity cues, which may reinforce the visibility and narratives of each identity category, as also suggested in previous research (e.g., Iori 2025). However, the equally frequent use of images featuring generic people without clear identity markers implies that this visual strategy is not consistent.

In light of the complementary aim of this research, namely to assess the sincerity or opportunism of corporate communication on LGBTQ+ support, the analysis yielded several insights. On the one hand, the predominance of communicative strategies focused on factual information - more frequent than explicit statements of commitment - may point to the companies' effort to highlight and express their social responsibility in more concrete terms, while also attempting to establish a connection and a sincere dialogue with LGBTQ+ stakeholders. This interpretation is reinforced by the seemingly complete, accurate, and relevant disclosure of inclusive policies and practices. On the other hand, representing LGBTQ+ stakeholders primarily as beneficiaries of corporate policies does not reflect their active role as contributors to inclusive corporate environments. This aspect, together with the presence of vague language, the relatively limited use of linguistic resources that convey clarity, and the use of visuals that reinforce a generic rather than identity-targeted approach, may indicate a form of communication that is partially constructed and potentially driven by opportunistic motives.

The findings of this study are based on a small exploratory corpus that still does not allow to draw clear conclusions on corporate sincere or opportunistic communication towards LGBTQ+ communities. In spite of this, the research has nonetheless provided linguistic, pragmatic, and multimodal analytical tools that could be used as a starting point to develop more systematic methodological toolkits to address this goal and that could be adopted for the analysis of larger and more homogeneous datasets, thereby enabling cross-sectoral and cross-cultural comparisons.

Bionote: Federico Zaupa, PhD, is a Research Grant Holder and Adjunct Professor at the University of Modena and Reggio Emilia. He holds a PhD in Human Sciences, with a thesis focusing on the discursive representation of LGBTQIA+ people and communities in the broadsheet press and press releases from LGBTQIA+ organisations in the UK and Italy, using the methodologies of corpus linguistics and critical discourse analysis. His research areas include language, gender, and sexuality in public discourse; Corporate Social Responsibility (CSR) communication; diversity, equity, and inclusion (DEI); and tourism

discourse. He is currently working on a University-funded project (FAR 2024) focusing on DEI and artificial intelligence in the discourse of sustainable tourism. He has also recently worked on a University-funded projects (FAR 2023; FAR 2022) and been involved in a National PRIN Project (PRIN 2020), titled “Communicating transparency: New trends in English-language corporate and institutional disclosure practices in intercultural settings.

Author’s address: federico.zaupa@unimore.it

Acknowledgements: This work was supported by the EU – NextGenerationEU, the University of Modena and Reggio Emilia, and Fondazione di Modena, with a research project entitled “THE MULTIFACETED COMMUNICATION OF DIVERSITY, EQUITY AND INCLUSION” (2023_PROGETTO_INTERDISCIPLINARE_2023_TURNBULL - CUP: E93C23002170005).

References

- Bondi M. and Sezzi A. 2025, *By Way of Example: The Transparency of Exemplification in Reporting Genres*, in “Iperstoria” 25, pp. 161-182.
- Bondi M. and Nocella J.J. 2023, *Building Trust in the Transport Sector during the Pandemic: A Cross-cultural Analysis*, in “Language and Dialogue” 13 [3], pp. 309-335.
- Breeze R. 2012, *Corporate Discourse*, Bloomsbury, London.
- Carroll A.B. 1979, *A Three-dimensional Conceptual Model of Corporate Performance*, in “Academy of Management Review” 4 [4], pp. 497-505.
- Castellani A. 2015, *Visual Aspects of CSR Reports: A Semiotic and Chronological Case Analysis*, in Duarte Melo A., Somerville I. and Gonçalves G. (ed.), *Organisational and Strategic Communication Research: European Perspectives II*, CECS-Centro de Estudos de Comunicação e Sociedade Universidade do Minho, Braga, pp. 129-149.
- Catenaccio P. 2008, *Press Releases as a Hybrid Genre: Addressing the Informative/Promotional Conundrum*, in “Pragmatics. Quarterly Publication of the International Pragmatics Association (IPrA)” 18 [1], pp. 9-31.
- Channel J. 1994, *Vague Language*, Oxford University Press, Oxford.
- Cheng W. and Warren M. 2003, *Indirectness, Inexplicitness and Vagueness Made Clearer*, in “Pragmatics. Quarterly Publication of the International Pragmatics Association (IPrA)” 13 [3], pp. 381-400.
- Cifalinò A. and Cutuli E. 2023, *Framing Diversity in Corporate Social Media Communication and Annual Reports*, in “Lingue Culture Mediazioni” 11 [1], pp. 89-116.
- Crawford Camiciottoli B. 2025, *Social Transparency in the Fashion World: Fashion Brands vs. Fashion Media*, in “Iperstori” 25, pp. 10-29.
- Dahlsrud A. 2008, *How Corporate Social Responsibility is Defined: An Analysis of 37 Definitions*, in “Corporate Social Responsibility and Environmental Management” 15 [1], pp. 1-13.
- Diani G. and Turnbull J.A. 2025a, *Investigating Transparency in Corporate Discourse about Safety*, in “Iperstoria” 25, pp. 183-203.
- Diani G. and Turnbull J.A. 2025b, *All Aboard: Diversity, Equity and Inclusion in the Sea Transport Sector*, in “International Journal of Linguistics” 17 [3], pp. 22-40.
- Falcone M. 2025, *Social Sustainability and Transparency: A Study of Person-first and Identity-first Language in Metropolitan Transportation Authority Communication*, in “International Journal of Linguistics” 17 [3], pp. 85-100.
- Fassin Y. and Buelens M. 2011, *The Hypocrisy-Sincerity Continuum in Corporate Communication and Decision Making: A Model of Corporate Social Responsibility and Business Ethics Practices*, in “Management Decision” 49 [4], pp. 586-600.
- Fombrun C.J. and Van Riel C.B.M. 2004, *Fame & Fortune: How Successful Companies Build Winning Reputations*. Pearson Education, Upper Saddle River.
- Fuoli M. 2018, *Building a Trustworthy Corporate Identity: A Corpus-based Analysis of Stance in Annual and Corporate Social Responsibility Reports*, in “Applied Linguistics” 39 [6], pp. 846-885.
- Garzone G.E. 2005, *Pragmatic and Discoursal Features of Annual Executive Letters: Observations on the Rhetorical and Evaluative Function of Concessive Constructions*, in Bondi M. and Maxwell N. (ed.), *Cross-Cultural Encounters: Linguistic Perspectives*, Officina Edizioni, Rome, pp. 130-143.
- Grice P. 1991, *Studies in the Way of Words*, Harvard University Press.

- Iori I. 2025, *Transparency on Social Media: A Case Study on Diversity and Inclusion on Transport for London's Instagram*, in "International Journal of Linguistics" 17 [3], pp. 101-119.
- Jin B. 2022, *A Corpus-Assisted Study of Vague Language in Corporate Responsibility Reports of the Cosmetics Industry*, in "Ibérica" 43, pp. 77-102.
- Kilgarriff A., Baisa V., Bušta J., Jakubíček M., Kovář V., Michelfeit J., Rychlý P. and Suchomel V. 2014, *The Sketch Engine: Ten Years On*, in "Lexicography" 1, pp. 7-36.
- Kim Y. 2023, *Consumer Responses Toward LGBTQ+ Diversity CSR: What Differences Do CSR Levels and CSR Fit Make?*, in "Corporate Communications: An International Journal" 28 [1], pp. 118-134.
- Kim H. and Lee T.H. 2018, *Strategic CSR Communication: A Moderating Role of Transparency in Trust Building*, in "International Journal of Strategic Communication" 12 [2], pp. 107-124.
- Koester A. 2007, *'About Twelve Thousand or So': Vagueness in North American and UK Offices*, in Cutting J. (ed.), *Vague Language Explored*, Palgrave Macmillan, London, pp. 40-61.
- Lischinsky A. 2015, *What Is the Environment Doing in My Report? Analyzing the Environment-As-Stakeholder Thesis Through Corpus Linguistics*, in "Environmental Communication" 9 [4], pp. 539-559.
- Malavasi D. 2010, *The Multifaceted Nature of Banks' Annual Reports as Informative, Promotional and Corporate Communication Practices*, in Evangelisti Allori P. and Garzone G.E. (ed.), *Discourse, Identities and Genres in Corporate Communication. Sponsorship, Advertising and Organizational Communication*, Peter Lang, Bern, pp. 211-233.
- Malavasi D. 2011, *'Doing Well by Doing Good': A Comparative Analysis of Nokia's and Ericsson's Corporate Social Responsibility Reports*, in Garzone G.E. and Gotti M. (ed.), *Discourse, Communication and the Enterprise: Genres and Trends*, Peter Lang, Lausanne, pp. 193-212.
- Malavasi D. 2017a, *From Corporate Social Responsibility Reports to Sustainability Webpages: An Analysis of Convergences and Divergences*, in Canepari M., Manfield G. and Poppi F. (ed.), *The Many Facets of Remediation in Language Studies*, LAP Lambert Academic Publishing, Riga, pp. 23-36.
- Malavasi D. 2017b, *Building Trust Through Corporate Identity: An Analysis of CSR Reports and Webpages*, in Salvi R. and Turnbull J.A. (ed.), *The Discursive Construal of Trust in the Dynamics of Knowledge Diffusion*, Cambridge Scholars Publishing, Newcastle upon Tyne, pp. 131-153.
- Malavasi D. 2023, *The Discursive Construction of Equality, Diversity and Inclusion: Insights from an Analysis of CSR Reports in the USA, UK and Japan*, in "Lingue e Linguaggi" 58, pp. 153-171.
- Malavasi D. 2025, *Diversity, Inclusion and Transparency in Rail Companies' Communication*, in "International Journal of Linguistics" 17 [3], pp. 41-55.
- Malavasi D. and Nocella J.J. 2025, *Doing the Locomotion: Aspects of Trust and Transparency in Railway Communication*, in "Iperstoria" 25, pp. 135-160.
- Nocella J.J. 2023, *Building Inclusiveness and Representing Diversity in Transportation Companies*, in Hofer-Bonfim et al. (ed.), *Diversity and Inclusion Across Languages*, Frank & Timme, Berlin, pp. 127-147.s
- Nocella J.J. 2025, *Transparency and Vague Language: Diversity, Equity, and Inclusion in the Rail Sector*, in "International Journal of Linguistics" 17 [3], pp. 56-71.
- Poppi F. 2025, *Airlines' Emission Disclosures: The Fine Line Between Opportunity and Environmental Inaction*, in "Iperstoria" 25, pp. 116-134.

- Searle J.R. 1969, *Speech Acts*, Cambridge University Press, Cambridge.
- Singh V. and Point S. 2006, *(Re)Presentations of Gender and Ethnicity in Diversity Statements on European Company Websites*, in “Journal of Business Ethics” 68 [4], pp. 363-379.
- Schnackenberg A.K. and Tomlinson E.C. 2016, *Organizational Transparency: A New Perspective on Managing Trust in Organization-Stakeholder Relationships*, in “Journal of Management” 42 [7], pp. 1784-1810.
- Sinclair J.M. 2004, *Trust the Text: Language, Corpus and Discourse*, Routledge, London.
- Szyndlar M. and Wąsikiewicz-Firlej E. 2019, *The LGBT Community as a Stakeholder in Communicating Corporate Social Responsibility: An Analysis of Selected Case Studies*, in “Scripta Neophilologica Posnaniensia” 19, pp. 191-221.
- Tognini-Bonelli E. 2001, *Corpus Linguistics at Work*, John Benjamins, Amsterdam.
- Turnbull J.A. 2023, *The Discursive Construction of Diversity & Inclusion in Corporate Websites*, in Hofer-Bonfim et al. (ed.), *Diversity and Inclusion Across Languages*, Frank & Timme, Berlin, pp. 149-168.
- Vaccaro A. and Madsen P. 2006, *Firm Information Transparency: Ethical Questions in the Information Age*, in *Social Informatics: An Information Society for All? In Remembrance of Rob Kling: Proceedings of the Seventh International Conference on Human Choice and Computers (HCC7), IFIP TC 9, Maribor, Slovenia, September 21-23*, Springer, US.
- van Leeuwen T. 2008, *Discourse and Practice: New Tools for Critical Discourse Analysis*, Oxford University Press, Oxford.
- Winkler B. 2000, *Which Kind of Transparency? On the Need for Clarity in Monetary Policymaking*, in “European Central Bank Working Papers” 26, European Central Bank, Frankfurt am Main.
- Yu D. and Bondi M. 2017, *The Generic Structure of CSR Reports in Italian, Chinese, and English: A Corpus-Based Analysis*, in “IEEE Transactions on Professional Communication” 60 [3], pp. 273-291.

© 2026 University of Salento - Coordinamento SIBA



<https://ese-journals.unisalento.it/>