Lingue e Linguaggi
Lingue Linguaggi 74 (2026), 65-85
ISSN 2239-0367, e-ISSN 2239-0359
DOI 10.1285/i22390359v74p65-85
https://ese-journals.unisalento.it, © 2026 Università del Salento

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 – Austin[1] 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 viewers[2] – 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:

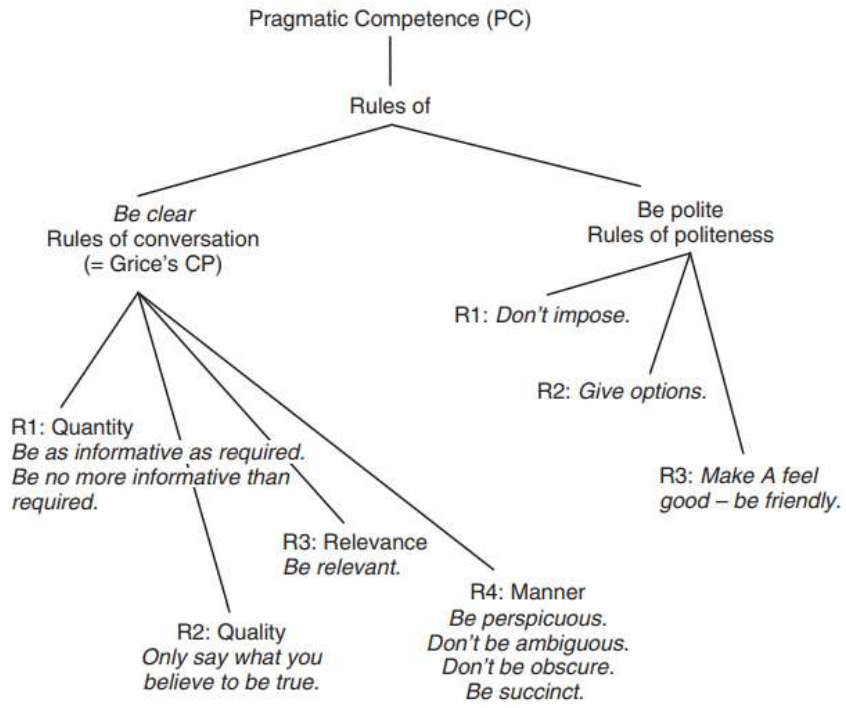

Figure 4
Lakoff's scheme of pragmatic competence (Watts 2003, p. 60; Flowerdew 2013, p. 107).

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:

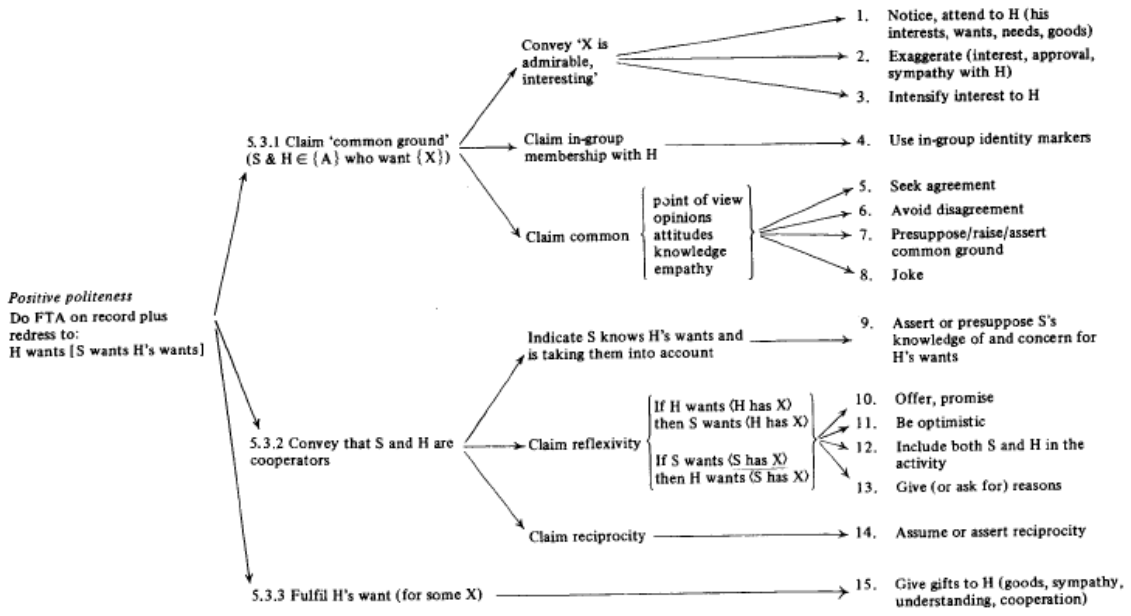

Figure 5
Brown and Levinson's chart of positive politeness strategies (1987, p. 102).

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 impositive 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 framework[4] 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., *rɪˈlæ::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. 2[5] (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., Irrmischer 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 Leeuween 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 Timişoara: Transactions on Modern Languages" 14 [1], pp. 23-30.

Streeck J. 2009, *Gesturecraft: The Manu-facture 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-meditationnear-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.