

Introduction

Jarret Geenen*

Introduction: multimodal (inter)action analysis

<https://doi.org/10.1515/mc-2023-0010>

Keywords: interaction analysis; multimodal analysis; multimodal interaction analysis; multimodal mediated theory

In the past 20 years, the multimodal turn in linguistics and the language sciences has properly gained momentum. While we can trace the beginnings of a multimodal empirical sentiment to the late 19th and early 20th centuries, it was not until approximately a hundred years later that a multimodal turn truly came to fruition. While detailing academic historical trajectories with any objective accuracy is fraught with difficulties, some discussion of the past is often necessary to understand how it is that we arrived in the present. Embarking on such a venture, however, requires preemptively apologizing for both the sparsity of historical detail and more importantly, to the many pioneers of multimodal approaches whose work is not mentioned explicitly (Charles Goodwin as a great example). It is beyond the scope of this brief introduction to provide a full historical account of multimodality more generally and beyond its breadth to detail all the work of the many contributors to multimodal approaches to language and social interaction. The historical trajectory I sketch is thus, a significantly abridged one and specifically limited to better understand the emergence of the methodological framework that is the focus of this particular special issue; Multimodal (inter)action Analysis (Norris 2004, 2011, 2019, 2020).

While the empirical efforts of people like Birdwhistle (1955), Hall (1959), Scheflen (1964) and Kendon (1967) cannot be overstated, it is important to acknowledge that general ideas about the multimodality of language and social interaction predate this work quite considerably. Long before empirical exploits probing the relationships between language and other non-verbal behavior, philosophers of language, anthropologists and psychologists of the early 20th century already showed an acute awareness that language was simply one of many meaning making resources mobilized in communication. The meaning of non-verbal behavior was significant in Darwin's (Darwin 1872) work and already, in 1927, Edward Rowell was exploring the relationships of gestures and speech. Karl Buhler sums the perspective up nicely positing that "all concrete speech is in vital union with the rest of a person's meaningful behaviour; it is among actions and is itself an action" (1990: 61; 1934: 52).

While ideas about non-verbal meaning making predate their efforts quite considerably, the early empirical work of people like Ray Birdwhistle, Albert Scheflen and Adam Kendon investigating non-verbal modes of communication like gesture and posture provided ample evidence of their meaningful impact on human communicative activity. In linguistic circles, however, much of the work in what was at the time referred to as Kinesics, was largely dismissed and relegated to the same pile of scientific nonsense as the Mahrabian Myth. It was not until the experimental psychological work of McNeill (1992) on manual hand gestures and the social semiotic efforts of Gunther Kress and Theo van Leeuwen that the language sciences truly began to acknowledge the value of non-linguistic phenomena in the communicative equation. The latter, Kress and Van Leeuwen's (1996) now seminal *Reading Images: The grammar of visual design*, had a profound impact on systemic functional linguistics and applied linguistics more broadly. Their work signaled the first truly systematic investigation of non-verbal (albeit primarily textual) meaning making, providing a vast inventory of analytical tools to be used in social semiotic approaches to visual meaning-making. There was, however, a distinct limitation to their approach, and that was the lack of direct applicability to real face-to-face human social interaction.

*Corresponding author: Jarret Geenen, Department of Modern Languages and Culture, Radboud Universiteit Nijmegen, Comeniuslaan 4, Nijmegen, 6500 HC, Netherlands, E-mail: jarret.geenen@ru.nl

As a sociolinguist and discourse analyst with an empirical focus on language and identity production in social interaction, Sigrid Norris recognized the potential value of considering non-verbal phenomenon in the communicative equation. The growing popularity of Kress and Van Leeuwen's (1996) work along with its lack of applicability to social interaction directly, provided a motivation to explore multimodal meaning-making in face-to-face contexts informed by Scollon's (1998) Mediated Discourse Theory. Combining Scollon's (1998) focus on mediation with insights from earlier exploits in kinesics, Norris built an analytical framework now known as Multimodal (Inter)action Analysis (2004, 2011, 2019, 2020), which could be systematically applied to face-to-face human social interaction. The framework which facilitates a fine-grained analysis of non-verbal and verbal communicative behavior has become a staple in applied linguistics and has proven to be a supremely useful means for probing multimodal relationships in the pragmatics of social interaction. Paramount in this approach, is the utility of the notion of mediation and the analytical unit of the mediated action which we can trace back to the work of James Wertsch's (1991) and his Mediated Action Theory.

Inspired by the paradigm shifting ideas and approaches of Russian psychologist Lev Vygotsky, James Wertsch was very much taken by the utility of the notion of mediation as it pertained to both human psychological development and human social action. Instrumental in bringing Vygotsky's anti-reductionist and socio-cultural historical approach to human psychology to the masses (Wertsch 1985), the socio-historical perspective continued to permeate his own work much into the future. Vygotsky's (Vygotsky 1978) perspective and approach greatly influenced the formation of what would come to be known as Wertsch's Mediated Action Theory (1991, 1998) which held the notions of human action as well as psychological and technological mediation as paramount. Central in Wertsch's work was the notion of mediated action which he argued was the single most useful ecological unit of analysis on the basis that, as a discrete unit, it maintained as much socio-cultural, historical and institutional complexity as is possible. The mediated action refers to the social actor, acting with or through mediational means or cultural tools. As such, the action-based unit always and irreducibly has, individual, socio-cultural, historical and institutional trajectories permeating its nature.

The insight that all human action is mediated by mediational means and/or cultural tools has wide reaching ramifications. The analytical repercussions are that the influence of mediational means and/or cultural tools which mediate action must be recognized as consequentially shaping and influencing such action. As such, traditional notions of individual agency require abandoning and become socio-historically distributed. Agency in action must be seen as shaped and coerced through the cultural tools which mediate such action. Additionally, recognizing all actions as mediated by mediational means or cultural tools requires acknowledging that all actions are inevitably social by nature. The material nature of any action is consequentially shaped by the structure, organization and development of psychological and/or material cultural tools. The consequential influence of these tools means that the socio-cultural, historical and institutional are always ever-present in all forms of mediated action.

In the late 90s, Scollon (1998) recognized the potential in Wertsch's (1991, 1998) unit of analysis and incorporated its core tenants in his own approach to contemporary sociolinguistics in his Mediated Discourse Theory and Mediated Discourse Analysis. Scollon (1998) argued that discourse and discursive activity is best conceived of as social action and that all social action should be conceptualized as mediated action. Scollon posited that understanding language as it operates in social life required approaching language not as an abstract semiotic system but instead as a tool through which humans undertake action. While Scollon (1998) implicitly acknowledged the potential applicability of the mediated action as a unit of analysis for non-verbal and non-textual forms of communicative phenomenon, Norris (2004) took on this challenge more directly in the development of Multimodal (Inter)action Analysis as an analytical framework.

While Wertsch (1991, 1998) and Scollon (1998, 2001) make a compelling case for the theoretical utility of the mediated action as a unit of analysis, the concrete application of the unit for the analysis of multimodal phenomena posed numerous challenges. First, acknowledging that verbal actions tend to occur with additional forms of non-verbal bodily actions requires determining the scope of what can be identified as a mediated action. In some cases, this might result in the mediated action having certain properties (perhaps involving two communicative modes) and in others, having very different properties (involving four communicative modes and additionally, various technological tools). This variability is obviously undesirable and would result in a difficult

consistent application of the unit itself in analysis. Norris managed to overcome this hurdle (and a few others) by positing the utility of distinguishing mediated actions across individual modes of communication.

Partitioning of the mediated action into lower-level mediated actions and higher-level mediated actions with lower-level actions being the primary unit of analysis managed to overcome two equally problematic issues. As mentioned, ambiguity in definition is overcome because lower-level actions are relevant to individual modes. At any given moment, many lower-level actions may (or may not) temporally coincide. Additionally, by defining a lower-level action as the smallest pragmatic meaning unit of any individual communicative mode helps overcome the analytical issue associated with the differentiated material and structural organization of individual communicative modes.

One of the primary hurdles facing a multimodal analytical framework is determining a single unit of analysis and this is a result of the very different materialities and organizational properties of modes of communication. For those interested in human communication rather than abstract properties of individual linguistic systems, an utterance may be treated as a perfectly suitable unit of analysis. However, the properties of an utterance like audibility, mediated by a natural human language and fleeting material longevity may not be present in other modes of communication. Gesture, for instance, is not really audible in a traditional sense and is typically also not linear and compositional (cannot add gestures together to make a new gesture in the same way as with individual words). Gesture, as a mode, tends to exhibit material and organizational properties that are very different to language, and therefore, it is very hard to define an *utterance* in gesture. However, it has been quite successfully argued and it is now generally empirically accepted that the smallest pragmatically meaningful unit of gesture is a complete gesture which includes, at a minimum, the stroke or stroke hold phase of the gesture (McNiell 1992). In defining a lower-level action as the smallest pragmatic meaning unit of any individual communicative mode, differences in materiality and organizational structure are accommodated and temporal multiplicity is also perfectly acceptable all while maintaining a single unified unit of analysis. The unit is analytically applicable across communicative modes without ambiguity and issues in consistent application. This is accomplished while also maintaining the theoretical utility of recognizing all actions as mediated actions.

In addition to being flexible enough in definition to be applied across the modal spectrum, there is another distinct advantage; the unit is adaptable and amenable to empirical insights generated in any single modal domain. The multimodal turn in applied linguistics in the late 90s has been followed more recently with prolific increases of interest in areas like second language acquisition, experimental linguistics, psycholinguistics, cognitive psychology and cognitive neuroscience. Many of these disciplines are quite traditionally positivist in their methodological leanings and as a result, controlled experimentation allows for an in-depth look at specific components of individual communicative modes. The knowledge generated in these fields greatly enrich Multimodal (Inter)action Analysis. Findings can inform refinement in the analysis of individual modes and more importantly, deeper understandings of the communicative function of any single mode can significantly inform investigations and conceptualizations of the cross-modal and intermodal relationships which are paramount in real-time social interactions.

In maintaining an empirical and analytical focus on social action as mediated and by employing an analytical unit which is consistent while flexibly sensitive across various modes of communication, Multimodal (Inter)action Analysis is an analytical framework which is particularly advantageous given the evolving dynamics of both the communication sciences and the ecological landscape of human communication. First, I would like to consider the latter in relation to the analytical focus on mediation as paramount.

One of the central theoretical tenants which Multimodal (Inter)action Analysis (MIA) has adopted from its discursive and psychological antecedent approaches is the centrality of mediation in all forms of social action. The analytical advantages of a focus on mediation are multiple and make MIA particularly valuable in the changing socio-technical ecology of our ever-changing world. Not only does this focus on mediation distinguish MIA from other social action-oriented approaches to the study of human communication, it also makes the analytical framework perfectly suited to tackle some of the empirical puzzles which have been created by the increasing permeation of technology in our communicative lives. While the global spread of Covid-19 made computer mediated forms of social interaction a necessity for not only work-life but also social lives, the technology which came to the fore during this time had long preceded its global necessity. Various technologies (hardware and

software) are now ubiquitous in our personal and work lives. Much of our social and organizational communicative activities throughout the day involve social messaging or video-conferencing applications.

Multimodal (Inter)action analysis has already proved quite valuable in unravelling some of the complexities which emerge through new technologically-mediated forms of social interaction. In the fractured interactional ecologies which emerge when communication is complexly mediated by new technologies, non-verbal modes take on a new importance and appear to be pivotal for exemplifying divergence in stance or disagreement (Norris and Pirini 2017), can be vital for the production of identity (Geenen 2017) and can be exploited to help facilitate smoother interactions with pre-verbal children and toddlers (Geenen 2018, 2020). Other work has shown how non-verbal actions can be pivotal to accomplishing collaborative tasks via video-conferencing technology (Geenen et al. 2021) but that the distribution of attention and the interactional demands can also result in miscommunication and misunderstanding (Norris and Geenen 2022). Elsewhere, Multimodal (inter)action Analysis has been applied to practices of interpreter training (Krystallidou 2014) and the training of teachers (Christensson 2020). In all of this work, the application of Multimodal (inter)action Analysis with a demonstrable focus on forms of mediation and non-verbal modes of communication has revealed interactional complexities which might otherwise be missed with language-centric methodologies or without taking seriously the consequentiality of technological tools and their mediating of communicative actions.

Importantly, the utility of a multimodal analytical methodology for dissecting the complexities of technologically mediated forms of social (inter)action is only growing. One obvious factor contributing to this utility is the increasingly multimodal turn in the language and communication sciences. Now, more than ever before, taking serious the multimodality of all social interaction requires an analytical methodology which is suitable for the task. While there have been dramatic increases in efforts to incorporate multimodal phenomenon into other language and discourse-based analytical methods, there remain stubborn issues with transcription, analysis and how epistemologies and ontologies underlying these approaches implicitly value language first and non-verbal modes only in an additional form. Multimodal (Inter)action Analysis, alternatively, maintains the multimodality of social interaction at its core without overtly prioritizing any mode *a priori* or in isolation. A standardized visual transcription method and analytical protocol further enhance its applicability across an array of social science domains and disciplines. Of equal importance in our ever-changing technological environment is the utility of the mediated perspective for dealing with the changing communicative ecological landscape.

One empirical benefit of the mediated approach is that mediation is at the analytical core of the framework. One immediate consequence of this is that the technological tools through which action is taken are always an inextricable component of the nature of that action. In our contemporary technological landscape, acknowledging that technological tools shape mediated actions in consequential ways is supremely important. As touched upon earlier, traditional notions of unified agency or individuated and uncomplicated identity production in communicative action must be abandoned in favour of recognizing the results of this mediation. While this has always been analytically important, technological advancements and increasingly technologically mediated social interaction make it paramount.

The classroom is one domain which is increasingly permeated by intersecting forms of technological and material mediation. Two articles in this special issue take up practices in the classroom specifically, probing the multimodal nature of teaching and learning practices. Bernard-Mecho (this issue) investigates forms of meta-discourse which are pervasive in university lectures and finds that this meta-discourse serves both active and passive purposes. While it can be used organizationally as a means to help guide and steer the audience during a lecture, it can also be used in the background as a type of filler. Bernard-Mecho highlights the need to raise awareness about multimodal performance and multimodal literacy as it pertains to the training of teachers and additionally highlights the importance of multimodal genres.

Mejia-Leguna (this issue) applies MIA to Critical Learning Episodes in the EFL classroom investigating the multimodal coordination of gaze, posture, head-movement and speech in moments where learning is recognized as explicitly taking place or as inhibited. The analysis suggests that non-verbal modes do not just play an ancillary role to language in the language learning classroom, but rather, play central roles in many of the practices which emerge. Mejia-Leguna shows how modes like gesture, posture and gaze can contribute to enhancing aspects of shared attention and are used quite demonstrably for understanding confirmation checks. The findings of the

study highlight the consequentiality of non-verbal modes in teaching and learning practices, even when the object of the learning is language acquisition.

Multimodal (inter)action Analysis originally developed through an investigation of language and identity production (Norris 2002) where it became clear that identity producing actions were not uniquely verbal. Social actors can and often do undertake identity-producing actions through modes other than spoken language and multimodal analysis is required to understand how differentiated identity elements can be produced simultaneously (Norris 2004). The empirical focus on identity producing actions which influenced the emergence of the framework in the first place, is carried on in this special issue by Matelau & Sagapolutele (this issue) and Rajic (this issue) who take up the complexities of identity production through multimodal ensembles explicitly.

Matelau & Sagapolutele investigate the production of Samoan identity in New Zealand which is complexly influenced and shaped by social, historical and institutional factors. Through the analysis of a creative practice, the authors detail identity production through many layers of discourse and highlight overlapping and differing features of identity production across sites of engagement. Additionally, the authors posit that dance as a creative practice is a site of identity negotiation with complexly intersection individual, interpersonal, cultural and political factors at play.

Rajic's visual essay explores the differences in non-verbal behavior and identity production comparing Serbian and English native speakers in New Zealand. The results suggest that there are systematic differences in non-verbal expressiveness between Serbian-English bilingual speakers and New Zealand English monolingual speakers. The results have significant repercussions for cross-cultural studies of non-verbal behavior and identity production. This further highlights the need for a better understanding of commonalities and differences cross-culturally in our every expanding, multicultural and multilingual communicative ecology.

As an analytical framework which has increasing utility as the multimodal orientation of the language sciences increases and as contemporary mediated forms of social interaction are on the rise, Multimodal (inter)action Analysis will surely have a place in our future. Given the flexibility of the framework itself, its strong theoretical underpinnings and its ability to accommodate for insights across a diversity of communications and psychological disciplines, widespread use of the framework will inevitably result. This has a myriad of advantages as with adoption, inventories of analytical tools grow and importantly, our understanding of modal interrelationships expands facilitating more insightful analyses the pragmatics of social interaction across all domains of social life.

References

- Birdwhistle, R. (1955). Background to kinesics. *A Rev. Gen. Sem.* 13: 10–18.
- Buhler, K. (1990). *Theory of language: the representational function of language*. Translation by Goodwin D. F. John Benjamins, Amsterdam.
- Christensson, J. (2020). Interactional role shift as communicative project in student teachers' oral presentations. *Multimodal Commun.* 9: 1–16.
- Darwin, C. (1872). *The expression of the emotions in man and animals*. John Murray, London.
- Geenen, J. (2017). Show (and sometimes) tell: identity construction and the affordances of video-conferencing. *Multimodal Commun.* 6: 1–18.
- Geenen, J. (2018). Multimodal acquisition of interactive aptitudes: a microgenetic case study. *Pragmat. Soc.* 9: 519–546.
- Geenen, J. (2020). Objects and materiality in pragmatic development: here-and-now to then-and-there. *Multimodal Commun.* 10: 19–36.
- Geenen, J., Matelau-Doherty, T., and Norris, S. (2021). Visual transcription: a method to analyse the visual and visualise the audible in interaction. *RFMV* 5: 51–80.
- Hall, E. (1959). *The silent language*. Doubleday, New York, NY.
- Kendon, A. (1967). Some functions of gaze direction in social interaction. *Acta Psychol.* 26: 22–63.
- Kress, G. and van Leeuwen, T. (1996). *Reading images: the grammar of visual design*. Routledge, London, England.
- Krystallidou, D. (2014). Gaze and body orientation as an apparatus for patient inclusion into/exclusion from a patient-centred framework of communication. *Interpreter Transl. Train.* 8: 399–417.
- McNeill, D. (1992). *Hand and mind: what gestures reveal about thought*. University of Chicago Press, Chicago, IL.
- Norris, S. (2002). The implication of visual research for discourse analysis: transcription beyond language. *Vis. Commun.* 1: 97–121.
- Norris, S. (2004). *Analyzing multimodal interaction: a methodological framework*. Routledge, London.
- Norris, S. (2011). *Identity in (inter)action: introducing multimodal (inter)action analysis*. De Gruyter Mouton, Berlin and New York, NY.
- Norris, S. (2019). *Systematically working with multimodal data: research methods in multimodal discourse analysis*. Blackwell-Wiley, Hoboken, NJ.
- Norris, S. (2020). *Multimodal theory and methodology: For the analysis of (Inter)action and identity*. Routledge, Abingdon, UK.

- Norris, S. and Geenen, J. (2022). Intercultural teamwork via videoconferencing technology. A multimodal (Inter)action analysis. In: Kecskes, I. (Ed.). *Cambridge handbook of intercultural pragmatics*. Cambridge University Press, Cambridge, pp. 552–587.
- Norris, S. and Pirini, J. (2017). Communicating knowledge, getting attention, and negotiating disagreement via videoconferencing technology: a multimodal analysis. *J. Organ. Knowl. Commun.* 3: 23–48.
- Rowell, E.Z. (1927). Gesture – an exceptional use. *Am. Speech* 3: 38.
- Schefflen, A. (1964). The significance of posture in communication systems. *Psychiatry* 27: 316–331.
- Scollon, R. (1998). *Mediated discourse as social interaction: a study of news discourse*. Addison, Wesley Longman Limited, London.
- Scollon, R. (2001). *Mediated discourse: the nexus of practice*. Routledge, London and New York, NY.
- Vygotsky, L.S. (1978). *Mind in society: the development of higher psychological processes*. Harvard University Press, Cambridge, MA.
- Wertsch, J.V. (1985). *Vygotsky and the social formation of mind*. Harvard University Press, Cambridge, MA.
- Wertsch, J.V. (1991). *Voices of the mind: a sociocultural approach to mediated action*. Harvard University Press, Cambridge, MA.
- Wertsch, J.V. (1998). *Mind as action*. Oxford University Press, New York, NY.