Essay

Ciara R. Wigham and Müge Satar*

Adapting and extending multimodal (inter)action analysis to investigate synchronous multimodal online language teaching

https://doi.org/10.1515/mc-2024-0048
Received April 29, 2024; accepted July 18, 2024; published online August 8, 2024

Abstract: Multimodal (inter)action analysis offers a powerful and robust methodology for the study of action and interaction between social actors, their environment, and the objects and tools within. Yet its implementation in the analysis of synchronous multimodal online data sets, e.g. (inter)actions via videoconferencing, is limited. Drawing on our research in understanding teacher-learner (inter)actions in instruction-giving fragments in synchronous multimodal online language lessons, we describe and illustrate the ways in which we adapted and extended some of the methodological and analytical tools. These include (1) the use of a grounded theory approach in delineating and identifying higher-level actions, (2) the embodiment and disembodiment of frozen actions, (3) electronic print mode, (4) semiotic lag, (5) semiotic (mis)alignment, (6) modal density (mis) alignment, and (7) how modal density can be achieved by brisk modal shifts in addition to through modal intensity and complexity. We conclude by a call for further educational research in online teaching platforms using the framework to have richer understandings of the (inter)actions between social actors with particular roles and identities (teachers-learners), their environment, and the objects and tools within, which bring their "own material properties, feel and techniques of use, affordances and limitations" (Chun, Dorothy, Richard Kern & Bryan Smith. 2016. Technology in language use, language teaching, and language learning. *The Modern Language Journal* 100. 64–80: 65).

Keywords: multimodal interaction analysis; synchronous online language teaching; semiotic (mis)alignment; modal density (mis)alignment; electronic print mode

1 Theoretical background of the methodology

Multimodal (Inter)action analysis (MIA) is receiving increasing attention since the multimodal turn in applied linguistics. This is evidenced by special collections, such as the one edited by Geenen (2023) in this journal positioning MIA as an analytical method for the future. MIA is a multimodal discourse approach and is complementary to systemic functional (Halliday 1978) and social semiotic (Bezemer and Jewitt 2009; Jewitt 2014) frameworks (Norris 2020). While the former explores the ways in which language is organised and used to accomplish a number of social functions (ideational, textual, and interpersonal meanings), the latter aims to understand the agency of social actors as well as social and power relations between them (Jewitt et al. 2016). Multimodal (inter)action analysis combines elements of multimodal discourse analysis and interactional sociolinguistics to unpack mediated actions. It has evolved from the fields of applied linguistics, anthropological linguistics, sociolinguistics, discourse analysis, and socio-cultural psychology, and is significantly influenced by social semiotic theories (Norris 2016).

***Corresponding author: Müge Satar**, Newcastle University, Newcastle upon Tyne, NE1 7RU, UK, E-mail: muge.satar@newcastle.ac.uk
Ciara R. Wigham, ACTé, Université Clermont Auvergne, Clermont-Ferrand, France, E-mail: ciara.wigham@uca.fr

Theoretically, multimodal (inter)action analysis is grounded in the idea that human actions are inherently linked with their environment and the objects, and thus offers explanatory and analytical tools for a fine-grained investigation of interconnections between social actors, material objects, and the world surrounding them (Norris 2016, 2020). This is conceptualised in the use of the word *(inter)action* (as opposed to *interaction*) to highlight that every action is potentially an interaction "that an individual produces with tools, the environment, and other individuals" (Norris 2011: 1).

With an emphasis on social action, two principles guide the analysis: all actions are communicative, and all actions have a history (Norris 2019). As such, social practices can be delineated as mediated actions with a history, emerging directly from the actions of the social actors (Norris 2020). Similar to multimodal discourse analysis (Scollon 1998), MIA focuses on human actions, all of which are mediated (Norris 2004, 2013, 2019, 2020). In applied linguistics, the concept of mediation comes from the sociocultural theory of learning which emphasises how learners' relationships are mediated by symbolic tools, like language (Lantolf 2000). In MIA, human interactions amongst themselves and with their physical, cultural, social, and psychological surroundings occur through mediational means, which influences their understanding and interpretation of the nature of the world and of their collaboration. Mediated action also illustrates the *unresolved dialectic* that occurs between social activity and the cultural products and semiotic resources that mediate it (Wertsch 1998) such as language, objects, technology, and practices. These are perceived as the *cultural tools* or *mediational means/methods*, each with unique affordances and limitations (Jones and Norris 2005), which need to be considered to understand human social behaviour while completing specific tasks.

Drawing on Wertsch's (1998) concept of mediated action, Norris (2004, 2016, 2019, 2020) proposes three units of analysis: lower-level mediated actions, higher-level mediated actions, and frozen actions. These analytical units enable the researcher to analyse (1) the multiplicity of (inter)actions that social actors perform (semi-)simultaneously (such as cooking while talking on the phone) and (2) the ways in which each social actor may perform and experience "a co-produced (inter)action differently" as each social actor's focus is not necessarily on the same (inter)actions at the same time (Norris 2020: 3).

Each social (inter)action is also potentially multimodal as "all modes together build one coherent system of communication" (Norris 2020: 2). A *mode* is then a "system of mediated action with regularities" (Norris 2013: 156). (Inter)actions can be produced through multiple modes being employed at the same time, with some becoming more relevant than others in and for an (inter)action. Through a micro-analytic investigation of how lower-level actions (which make-up a higher-level action) are multimodally configured, it is possible to identify the modal density of each higher-level (inter)action. This enables the researcher to position each higher-level action on an attention/awareness continuum depending on their relative modal density for each social actor. By doing so, we can discover "how attention and awareness levels vary in (inter)actions", and "how social actors co-producing an (inter)action pay different (or the same) focused attention to the (inter)action" (Norris 2020: 8).

While there are emotional, physical, and psychological states of attention, what is relevant here is (inter) actional attention on a foreground-background continuum to analyse the differentiated interactive attention levels in which social actors engage. (Inter)actions produced with a higher modal density are foregrounded, while those with a lower modal density are in the background of a person's attention/awareness. Norris (2004) describes modal density as being achieved either through modal intensity or modal complexity. A mode takes on high modal intensity when the higher-level action being performed by the social actor would not be possible if the mode had not been intensified. For example, a gesture towards a child to not interrupt when a parent is engaged in a phone conversation has high modal intensity because it is the only mode used to achieve the higher-level action of telling the child not to interrupt, and without which the action cannot be possible.

Modal complexity occurs when the modes that a social actor draws upon to construct a higher-level action are intricately intertwined, with no one mode taking on particularly high intensity nor a change in modes substantially altering the higher-level action. For example, two social actors wrapping birthday gifts where they employ object handling, gestures, gaze, and spoken language mode in an intertwined fashion has high modal density through modal complexity.

2 Studies on synchronous multimodal online interaction

Research in synchronous online interaction has gained traction particularly since the Covid-19 pandemic as isolation and social distancing measures forced social and educational interactions to move online. While the multimodal features enabling spoken, written, and visual interaction via screens may feel face-to-face (Develotte et al. 2011), synchronous online interaction platforms bring their "own material properties, feel and techniques of use, affordances and limitations" (Chun et al. 2016: 65). Research in the area pre-dates the pandemic. Early examples of studies investigating multimodal interaction in synchronous computer mediated communication have been published since the early 2000s (e.g. Chanier and Vetter 2006; Hampel and Baber 2003; Payne and Whitney 2002). Within online language teaching, learning, and intercultural communication settings, various methods have been used to illuminate multimodal aspects of pedagogical interactions, such as multimodal discourse analysis (e.g. Lee et al. 2019), conversation analysis (e.g. Cappellini and Azaoui 2017), and social semiotics (e.g. Satar et al. 2023).

While the aforementioned discursive, multimodal, and ethnomethodological approaches are useful, they fail to describe the use of language along with other modes as part of actions individuals perform alone, with others, and/or in interaction with their environment acting in and with the socio-cultural world. Our interest in employing MIA for the study of language lessons in online synchronous multimodal communication stems partly from this unique positionality of the method as well as the need for "researching online language learning … from new and innovative approaches, … [which] requires a conscious effort and redirection of research energies to deal with the material differences that make online language learning unique" (Stickler and Hampel 2019: 24).

There are only a few studies that have implemented MIA to investigate actions and interactions within the intersection of in-person and online communication via videoconferencing. For instance, Norris (2016) explored "families (inter)acting with family members via skype or facetime across the globe" (p. 141). Through notions of mediation, modal density, and attention/awareness continuum, she demonstrated how a social actor's attention shifted between three higher-level actions of *engaging with research project* in the physical environment, i.e. his own house, *Skyping with family members in Australia* on a laptop, and *interacting with his girlfriend* who was sometimes present in the same room. This set-up required two types of recordings (screen-recording on the laptop, and an external camera that records an in-room view) and two researchers observing the interaction and taking notes.

Geenen (2017) also investigated family interactions via videoconferencing using the same dataset reported in Norris (2016), and focused on the *showing of objects, entities, and artefacts*, particularly the ways in which this contributed to young children's agentive identity formation. Investigating *showing* as an interactive move, he elucidated how social actors explicitly acknowledged "the relationship between the objects, their interactive relevance and the frozen actions embedded in them" (p. 13). Norris and Makboon (2015) had shown how identity markers were available as frozen actions in the print mode and objects backgrounded in social actors' attention/awareness continuum in in-person interaction. Geenen (2017) extended the notion of frozen actions by demonstrating how identity markers were embedded and articulated in the actions of *showing* and *telling* as foregrounded in the social actors' attention/awareness.

Norris and Pirini (2016) investigated everyday knowledge communication, specifically the higher-level actions of *acknowledging knowledge*, *coordinating attention*, and *negotiating disagreement* in dyadic teamwork via videoconferencing. Two research participants were placed in different rooms, given physical materials (instructions, cut-outs of flowers in different colours, a half-drawn garden map, and pens) and asked to complete the garden model through negotiation of different aspects of the garden while interacting online via Skype. Similar to Norris (2016), interactions between the social actors were recorded through screen-recording on the laptops, while social actors' actions and interactions with the physical objects in the environment were recorded with an external video camera. Norris and Pirini (2016) demonstrated that the social actors performed different higher-level actions through the modes of gaze, gesture, posture, and object handling, with or without

language. *Accepting knowledge* was observed to be produced initially multimodally and its production in the mode of spoken language followed its non-verbal production.

Analysing the same pairs' interaction from the same corpus as Norris and Pirini (2016) and Geenen and Pirini (2021) demonstrated how the distribution of gaze patterns emerged with other modes in multimodal ensembles and were responsive to "the material and communicative exigencies of the higher-level action, which is in the foreground of a social actor's attention/awareness" (p. 99). Here the higher-level actions each actor engaged in shifted between communicating with the other actor and (inter)acting with the physical resources to complete the task.

Finally, investigating a corpus of dyadic teamwork via videoconferencing, Norris and Geenen (2022) explored the reasons why misunderstandings emerged. In this study, the participants interacted online via Skype while being physically present in different offices in the same building. They were given a role-play task which did not require any physical materials (except a task sheet with instructions) for task completion. The role-play involved pretending to stay at a given hotel and searching together for a restaurant for dinner. The authors showed how *interactive misalignment* emerged "due to a divergence in the ongoing practices" (p. 574) and participants "making different assumptions about each other" (p. 586). For example, while one was looking up hotel reviews, the other was searching for nearby restaurants on an online map. They evidenced how each social actor was individually engaged with a different higher-level action on their own screen (which was not shared), concluding that each social actor produced their own mediated actions rather than co-constructing the actions, and that misunderstandings between the participants were due to either participants' use of different practices for task completion or lack of a focus on the same higher-level actions at the same time.

While these studies shed light on the employment of various modes to coordinate (inter)actions taking place either online and sometimes in interaction with other actors and objects in the physical environment, they do not focus on the unique affordances of and challenges caused by the synchronous online environment in operationalising multimodal (inter)action analysis.

3 Identifying the methodological gap

In this article, we describe some of the challenges we experienced in investigating multiple higher- and lower-level actions in synchronous multimodal online lessons between three social actors (a language teacher and two learners), and how we adapted and extended some of the analytical tools of MIA in response to these challenges (Satar and Wigham 2020, 2023; Wigham and Satar 2021). These were required because the context differed in multiple ways from the previous work reported above.

First, none of the aforementioned studies on synchronous multimodal online interaction aimed to produce a comprehensive framework of higher-level actions which comprise a specific macro-level social action/practice. In our context, this was *giving instructions*. To achieve this goal, we required a more rigorous method of identifying and delineating higher-level actions grounded in the data.

Second, compared to Norris (2016) and Geenen (2017), the social actors in our work were alone in their physical spaces, which meant that there were no interactions with other actors beyond the screen.

Third, compared to Norris and Pirini (2016), Geenen and Pirini (2021), and Norris and Geenen (2022), all the resources used were electronic: the social actors did not (inter)act with any physical objects or artefacts. Thus, compared to Norris and Makboon (2015) – who explored frozen actions in objects and print in the background of social actors' attention/awareness – and Geenen (2017) – who extrapolated the notion of frozen actions and how they were foregrounded in actors' attention/awareness through spoken reports while showing objects in family Skype interactions – the frozen actions in our work could only be observed electronically on the computer screen.

Fourth, data were collected in naturally occurring settings where all the participants were in different countries and did not have the technical equipment or skills to record the in-room view. This meant that observing shifts in gaze direction to signal a move between different higher-level actions, i.e. completing the task by interacting with the physical environment and communicating with the other person (as in Geenen and Pirini 2021), or foreground/backgrounding of these higher-level actions in social actors' attention/awareness (as in

Norris and Pirini 2016) was more difficult to identify. In our data set, gaze was almost always directed at the communicative and material exigencies on the screen. Alternations in gaze direction between communicative and material elements took place only on the screen and were sometimes signalled through minor shifts in posture direction or head movement. These led us to investigate differences in the modal configuration of each social actors' screen (site of engagement) to understand similarities and differences in modal density of each higher-level action and then subsequently be able to place them on the attention/awareness continuum of each social actor. As we did so, we discovered instances of (mis)alignment in the available modes and semiotic resources for each actor as well as occasions where there were (mis)alignments in modal density between the social actors.

Fifth, Norris and Geenen (2022) focused on misunderstandings during teamwork, which appeared to stem from different task completion practices and different higher-level actions foregrounded in each actor's attention/awareness observable in the different tasks in which they were individually engaged with on the internet browsers on their own screens. While the authors explored *interactional misalignment* which was not observable in the spoken language mode but evidenced in social actors' interactions with their environment in other modes, they did not analyse or compare the modal configuration and modal density of each site of engagement, nor the ways in which misalignment existed when there was no physical or visual co-presence (social actors' webcam videos displayed on the other actor's screen were covered by the browser window during task completion).

Finally, as participants were in different countries with different levels of internet access, speed, and bandwidth (unlike Geenen and Pirini 2021; Norris and Geenen 2022; Norris and Pirini 2016), time delay between the availability of semiotic resources for different participants played an important role in interactions. Although software licences were bought and shared, learners experienced technical problems and could not install and record their screens during the online language lessons via Skype. As an alternative, a researcher joined the calls and recorded a potential learner view.

Bearing these differences in mind, we now describe how we adapted and extended some of the analytical tools of multimodal (inter)action analysis in response to these contextual and methodological challenges.

4 Adapting and extending MIA to analyse videoconferencing interactions

The synchronous multimodal online dataset upon which our work draws was collected during a research project that investigated instruction-giving in language teaching-learning (Satar and Wigham 2020, 2022, 2023). The dataset comprises a teacher (Craig) and learners of English as a foreign language involved in an online language lesson via a videoconferencing platform. Two lessons are referred to in this article: primarily, a lesson in which Craig interacted with two learners: Didem and Eda. We also refer briefly to a one-to-one lesson by Craig with the learner we called Kuzey. During the lessons, the social actors were physically in geographically distant locations and each actor was alone in their physical space.

Extract 1 (see Appendix A) concerns the social actors giving and receiving instructions for the learning activity. The latter required each learner to access a different electronic resource. The interactions between the social actors were collected through screen recordings by both the teacher and a researcher who also connected to the videoconferencing platform, Skype. The dataset was collected in Spring 2018. Ethics approval was obtained from Newcastle University's ethics committee and all social actors (participants) gave informed consent.

Norris (2019) offers step-by-step guidance on conducting MIA for small, medium, and large research projects. The first step in analysis is the process of delineating higher-level mediated actions (HLAs) in a table and then consolidating them to prevent either exaggerating rare occurrences or minimizing frequently observed higher-level mediated actions within the dataset. However, the specific technique for achieving this is not elaborated upon. We needed a rigorous way of delineating all HLAs within the macro context of instruction-giving to address our research question: What higher-level actions comprise experienced online language teachers' task instructions-as-process? (Satar and Wigham 2020) This required systematic identification of the HLAs.

Meija-Laguna (2023), who similarly applied MIA to a language teaching context, albeit face-to-face not online, suggested recording higher-level actions in a table to allow tagging and colour-coding to identify categories of higher-level actions (large scale higher-level actions). Our first step in adapting and extending MIA drew on a similar bottom-up, grounded approach to data analysis but drew on Grounded Theory (Strauss and Corbin 1998). Our process involved reviewing screen recordings of the synchronous multimodal online data multiple times, identifying HLAs, and crafting descriptions for these actions. We utilised ELAN multimodal transcription software (Sloetjes and Wittenburg 2008) to annotate and categorise HLAs. During the open-coding stage, we grouped similar HLAs together, refining and expanding the categories of actions until no new categories emerged, in other words until theoretical saturation was reached. Subsequently, we actively sought variations both between and within the categories through constant comparison to ensure that all HLAs were mutually exclusive and to systematically discern their significance or frequency. We contend that this method provides a more rigorous approach to bundling HLAs, particularly when, like in our context, research has a macro-level theoretical aim to propose a comprehensive framework of HLAs that make up a specific broader HLA. A description of our methodology and framework of HLAs is detailed in Satar and Wigham (2020) and in chapter 6, Section 6.5 of Satar and Wigham (2023), both available online as open-access content.

Whilst Norris (2004) suggests print is a visual mode referring to "written texts, including the language, the medium, the typography, and the content … [and] images in the printed media" (p. 44), we contend that this definition needs to also encompass electronic print media to account, for example in Extract 1, for the use of text chat within the videoconferencing environment (frame 1), collaborative online documents as electronic teacher-learning resources (frame 4), and URLs (frame 5). The case study detailed in Satar and Wigham (2022) specifically focuses on the electronic print mode which we also discuss in Satar and Wigham (2023, chapter 3). In the latter, we examine how the electronic print mode is combined with the spoken language mode for bimodal instruction.

In terms of adapting MIA, Extract 1 exemplifies how electronic print mode is both a disembodied and embodied mode. In frame 4, Craig's focal point of attention turns to a disembodied electronic resource in the print mode – an online document displayed on the web browser which includes the task information for Eda produced by the researchers in a prior HLA *preparing the resource*. This resource entails the frozen action *adding a URL* accomplished by the teacher prior to the online lesson. Norris describes frozen actions as "usually higher-level actions which were performed by an individual or a group of people at an earlier time than the real moment of the interaction that is being analyzed" and which are "frozen in the material objects themselves" (2004: 13–14). We consider frozen actions to also cover actions entailed in electronic objects in the interactional setting. Thus, Craig embodies the print mode highlighting a URL in the electronic resource (frame 5) through the object handling mode. In frames 7–11, the print mode continues to be embodied as Craig pastes the copied URL from the resource sheet into the videoconferencing textchat (frame 7) and subsequently sends it (frames 12–13) to the shared interactional environment. The URL/resource becomes available to the learners in frame 14. The HLA of *sending and allocating the resource* is achieved through the modal aggregate of gaze, (electronic) print and object handling.

In Frames 20, 23 and 25, Craig demonstrates critical semiotic awareness (Guichon 2013) regarding the possibility for semiotic lag. We define semiotic lag as the time difference between the communication of a message by one social actor and its reception by another due to online transmission delay (Satar and Wigham 2023; Wigham and Satar 2021). For example, laughter communicated by a lower-level mediated action (LLA) in the mode of facial expression combined with a LLA in the spoken language mode may no longer form a modal aggregate when received by another social actor due to weak Internet connection or technical issues (e.g. microphone misfunctioning). The reception of the LLA by one social actor may be at a different moment within the interaction than its temporal position in the communicator's interaction space and attention/awareness. Furthermore, when several social actors participate in the site of engagement, they may not receive the actions at the same time either. Craig signals his recognition that the learner may have access to the electronic resource at a different moment within the interaction. He does this through utterances and silence in the spoken language mode, and changing the pace of the interaction to account for semiotic lag.

The HLA of *sending and allocating the resource,* also illustrates a second concept specific to mediated online communication: *semiotic (mis)alignment* (Satar and Wigham 2023; Wigham and Satar 2021, chapter 3). We define

semiotic (mis)alignment as referring to having the same (*semiotic alignment*) or differing (*semiotic misalignment*) levels of access to and availability of semiotic means for each social actor.

Indeed, we extended MIA in our work as we identified sources of semiotic misalignment which may fragment or distort the shared interactional space. The first source of *semiotic (mis)alignment* we explored related to social actors' use of different software and/or hardware configurations, for example, participants using different devices to connect to the site of engagement. Depending on the social actor's device, features of the online platform, such as layout, screen design or menu items may be presented differently. These differences in presentation may alter the multimodal composition of each actor's individual site of engagement. For instance, whilst it may be possible to see all social actors' webcam images on a computer screen, this may not be the case when connecting to the same videoconferencing platform from a mobile device with a smaller screen because it may only display the actor who is actively contributing in the spoken language mode. This may then lead to semiotic misalignment between the participants. Some platforms also allow social actors to modify their individual layout resulting in the shared interactional space being viewed differently.

In another example, from our project dataset, our analysis revealed that another teacher relied on the learner webcam image positions in her own layout mode to orchestrate spoken language, gesture, gaze, posture when allocating task roles to learners. However, the teacher did not have access to the students' screen layout or information as regard to where each actor's webcam image was positioned on the learners' screens.

Another example concerns the layout mode and access to textchat in the print mode. For some actors in our dataset, including the teacher Craig, the textchat window was open throughout the interaction and therefore messages sent were displayed as frozen actions to which the social actors could later refer (see Extract 1). For other social actors, e.g. the researcher, however, text chat messages were displayed over the interlocutor's webcam image for a short period of time but would then fade out of the site of engagement (see Figure 1). These examples illustrate that software/hardware configurations may affect whether the multimodal aggregates are available to all the social actors, as well as the pertinence of these, with potential for semiotic misalignment and a loss of a common "site of display" (Jones 2009: 115).

A second source of *semiotic (mis)alignment* relates to social actors having access to different resources that are not visible or accessible to others. For example, in Extract 1 frames 4–5, Craig accesses and, in the object handling mode, interacts with an electronic resource that is not yet available within the learners' site of engagement. In previous work (Satar and Wigham 2023; Wigham and Satar 2021), we described how semiotic misalignment is often communicated by changes in a social actor's modal configuration. Craig, for example, communicated changes in semiotic alignment by fewer gestures and less accentuated facial expressions and gaze shifts (Figure 2) which were sometimes combined with a LLA in the spoken language mode. However, the changes

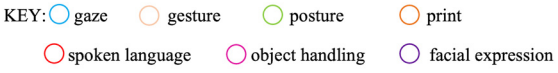


Figure 2: Teacher and Researcher access to textchat in the print mode
(Iteration 2, 00:22:14.850 and 00:22:05.070)

((Further conversation as regards Kuzey's response))

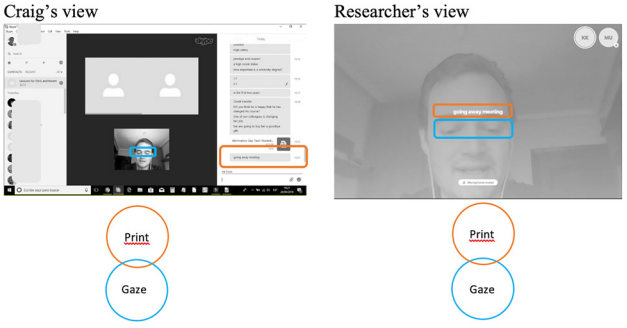


Figure 1: Teacher and researcher access to textchat in the print mode.

in modal configuration were not always apparent to the other social actor(s) who did not necessarily signal recognition of changes in Craig's attention/awareness unless there were explicit or significant shifts in the teacher's embodied modes.

As a result of semiotic misalignment, the modes that carry high or low modal density for each actor may differ resulting in instances of modal density misalignment. We discuss this concept in Satar and Wigham (2023, chapter 3). In comparison to semiotic misalignment which refers to social actors' access to different modes in their site of engagement, modal density misalignment refers to the differences in the LLAs or HLAs being foregrounded in each social actor's attention/awareness, regardless of the modes which are available to them in their site of engagement. For example, taking Extract 1, frame 5 (see Figure 2), as Craig utters "this one is for you", while the HLA of *copying the resource URL* is foregrounded in his attention/awareness (through print and

Figure 2: Craig's modal configurations. Due to technical difficulties for learners to record the interactions, here the researcher's screen recording is used as a proxy.

object handling modes), for any other social actor with the researcher view, the foregrounded HLA through the modes of spoken language and gaze is *being allocated a resource which is not yet visible*.

A final extension we propose to the concept of modal density is that, in addition to modal intensity and complexity, modal density can be achieved by brisk modal shifts with reference to speed and frequency. In Satar and Wigham (2023), we compared the modal configurations for different instantiations of the same HLA. While we found similar LLA modal configurations regarding intensity and complexity, a higher number of and more frequent modal shifts, for example the combination of brisk gaze shifts combined with frequent head nods, also led to higher modal density. Extract 5.4 is available online in Satar and Wigham (2022) and illustrates this concept which is discussed in Satar and Wigham (2023, chapter 5).

5 Conclusions

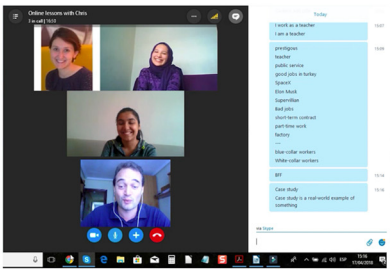
Whilst previous studies into online language learning and teaching have adopted interactional sociolinguistics, multimodal discourse analysis, conversation analysis, and semiotics for the macro or micro-analysis of talk-in-interaction, our application of MIA, with the mediated action as the fundamental unit of analysis, allows analysis (1) at both macro and micro levels through the exploration of higher- and lower-level mediated actions, and (2) of the hierarchical organisation of actions and modes, rather than only their sequentially. In applying MIA to synchronous multimodal online data, to better understand the interaction between social actors, their environment, and the object and tools within, we sought make methodological contributions to MIA.

First, by incorporating grounded theory to identify and delineate higher-level mediated actions more systematically, we ensured a more comprehensive and robust coverage of a specific macro-level social action/practice (*giving instructions*) observed within our dataset. Expanding the effective methodology of MIA to online interactions prompted us to broaden Norris' definition of the print mode to encompass electronic print, while also introducing several novel theoretical concepts tailored to synchronous online interactions: semiotic lag, semiotic (mis)alignment, and modal density (mis)alignment. We proposed the concept of semiotic lag to describe the desynchronisation of mode transmission, which can impact communication, the timing of actions, and how social actors engage with them. We argued that semiotic (mis)alignment can arise due to screen mediation, stemming from differences in semiotic meaning-making resources available to social actors, whether due to hardware or software variations affecting the layout mode, or disparities in resources accessed by different participants. Additionally, we proposed the idea of modal density (mis)alignment and suggested that modal density can be achieved through rapid shifts in addition to intensity and complexity of modes.

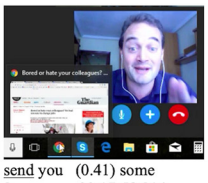
In summary, we hope these constructs contribute to the advancement of MIA methodology to enrich our understandings of mediated actions in online digital environments, offering insights into how communication can be affected by differences in the availability of and access to semiotic resources and the attentional focus of the social actors. We invite other colleagues working with MIA to engage with these concepts to test their robustness in datasets stemming from areas other than online language teaching and learning.

Acknowledgments: We are grateful to the learners and teachers who participated in this project. Full size image for Extract 1 are available at https://doi.org/10.25405/data.ncl.20315142, see Extract 4.3. For the purpose of open access, the author has applied a Creative Commons Attribution (CC BY) licence to any Author Accepted Manuscript version arising from this submission.
Research funding: We thank Newcastle University (Faculty of Humanities and Social Sciences Research Fund 2018 Spring call) and University Clermont Auvergne (Foreign researchers – short research visit 2018 call) for funding towards the project entitled "An Examination of Experienced Online Language Teachers' Multimodal Instruction-Giving Practices."

Appendix A: Extract 1.

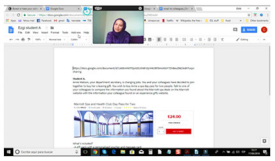


so what I'm gonna do is that I'm gonna
1 00:17:57.076

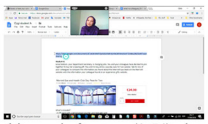


send you (0.41) some
2 00:17:58.214

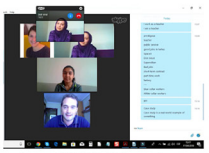
information (0.10) now (0.31)
3 00:17:59.895



Eda (0.41)
4 00:18:02.346



this one is for you (0.54) okay
5 00:18:02.912



u::hm (0.27) this one's
6 00:18:04.682

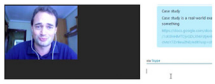
for Eda so
7 00:18:06.076

Didem
8 00:18:06.545

don't
9 00:18:06.831

look at this one yeah (0.05)
10 00:18:07.084

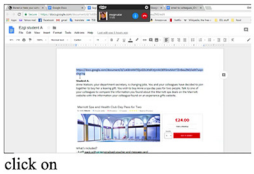
whatever you do
11 00:18:07.898



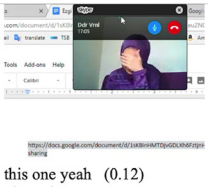
don't look at this one yeah (0.26)
12 00:18:08.708

Eda click on
13 00:18:09.655

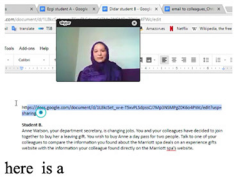
that one (0.34) uhm and Didem
14 00:18:10.662



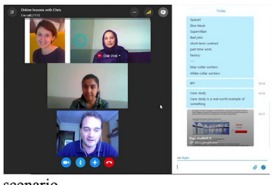
click on
15 00:18:12.932



this one yeah (0.12) uhm what we've got
16 00:18:12.762



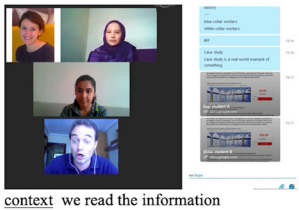
here is a
17 00:18:16.126



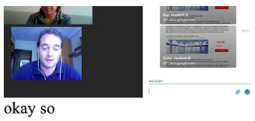
scenario
18 00:18:17.678

yeah (0.23)
19 00:18:18.399

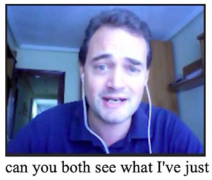
let me know when you can see it (0.20) and we'll read the
20 00:18:18.714



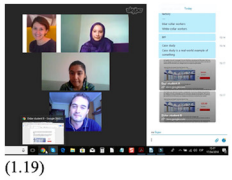
context we read the information together (0.23)
21 00:18:21.597



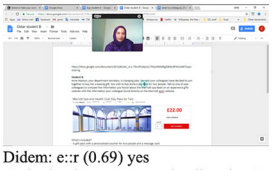
okay so
22 00:18:24.045



can you both see what I've just sent you?
23 00:18:24.920

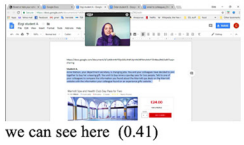


(1.19)
24 00:18:26.258

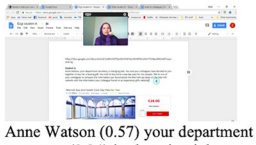


Didem: e::r (0.69) yes
Craig: it takes a moment loading the Google documents yeah (0.46)
okay here we go then (0.16)
25 00:18:28.135

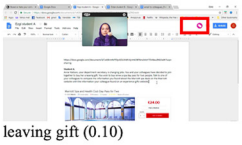
so at the top of the page (0.43)
26 00:18:34:434



we can see here (0.41)
27 00:18:36.461



Anne Watson (0.57) your department secretary (0.34) is changing jobs my goodness yeah (0.39) you and your colleagues have decided to join together to buy her a
28 00:18:37.823



leaving gift (0.10)
29 00:18::46.810

References

Bezemer, Jeff & Jewitt Carey. 2009. Social semiotics. *Handbook of pragmatics*, 1–13. Amsterdam: John Benjamins.

Cappellini, Marco & Brahim Azaoui. 2017. Sequences of normative evaluation in two telecollaboration projects: A comparative study of multimodal feedback through desktop videoconference. *Language Learning in Higher Education* 7(1). 55–80.

Chanier, Thierry & Anna Vetter. 2006. Multimodalité et expression en langue étrangère dans une plate-forme audio-sychnrone. *ALSIC* 9. 61–101.

Chun, Dorothy, Richard Kern & Bryan Smith. 2016. Technology in language use, language teaching, and language learning. *The Modern Language Journal* 100. 64–80.

Develotte, Christine, Kern Richard & Lamy Marie-Noëlle (eds.). 2011. *Décrire la conversation en ligne : le face à face distanciel*. Lyon: ENS éditions.

Geenen, Jarret G. 2017. Show and (sometimes) tell: Identity construction and the affordances of video-conferencing. *Multimodal Communication* 6(1). 1–18.

Geenen, Jarret G. (ed.). 2023. Special issue: Multimodal (inter)action analysis: An analytical framework for the future. *Multimodal Communication* 12(1).

Geenen, Jarret G. & Pirini Jesse. 2021. Interrelation. Gaze and multimodal ensembles. In Moschini Ilaria & Grazia Sindoni Maria (eds.), *Mediation and multimodal meaning making in digital environments*, 85–102. New York and London: Routledge.

Guichon, Nicolas. 2013. Une approche sémio-didactique de l'activité de l'enseignant de langue en ligne: réflexions méthodologiques. *Education & Didactique* 7(1). 101–115.

Halliday, Michael Alexander Kirkwood. 1978. *Language as social semiotic: The social interpretation of language and meaning*. London: Edwards Arnold.

Hampel, Regine & Baber Eric. 2003. Using internet-based audio-graphic and video conferencing for language teaching and learning. In Felix Uschi (ed.), *Language learning online: Towards best practice*, 171–191. Lisse: Swets & Zeitlinger.

Jewitt, Carey (ed.). 2014. *The Routledge handbook of multimodal analysis*. New York and London: Routledge.

Jewitt, Carey, Bezemer Jeff & O'Halloran Kay. 2016. *Introducing multimodality*. New York and London: Routledge.

Jones, Rodney H. 2009. Technology and sites of display. In Jewitt Carey (ed.), *The Routledge handbook of multimodal analysis*, 114–126. London: Routledge.

Jones, Rodney H. & Norris Sigrid (eds.). 2005. *Discourse in action: Introducing mediated discourse analysis*. New York and London: Routledge.

Lantolf, James P. 2000. *Sociocultural theory and second language learning*. Oxford: Oxford University Press.

Lee, Helen, Regine Hampel & Agnes Kukulska-Hulme. 2019. Gesture in speaking tasks beyond the classroom: An exploration of the multimodal negotiation of meaning via Skype videoconferencing on mobile devices. *System* 81. 26–38.

Meija-Laguna, Jorge Andrés. 2023. Classroom learning episodes in the EFL classroom: A multimodal (inter)action analytical perspective. *Multimodal Communication* 12(1). 23–44.

Norris, Sigrid. 2004. *Analyzing multimodal interaction*. New York and London: Routledge.

Norris, Sigrid. 2011. *Identity in (inter)action: Introducing multimodal (inter)action analysis*. Berlin: DeGruyter Mouton.

Norris, Sigrid. 2013. What is a mode? Smell, olfactory perception, and the notion of mode in multimodal mediated theory. *Journal Multimodal Communication* 2(2). 155–170.

Norris, Sigrid. 2016. Concepts in multimodal discourse analysis with examples from video conferencing. *Yearbook of the Poznan Linguistic Meeting* 2(1). 141–165.

Norris, Sigrid. 2019. *Systematically working with multimodal data: Research methods in multimodal discourse analysis*. Hoboken: Wiley Blackwell.

Norris, Sigrid. 2020. *Multimodal theory and methodology: For the analysis of (inter) action and identity*. New York and London: Routledge.

Norris, Sigrid & Geenen Jarret G. 2022. Intercultural teamwork via videoconferencing technology: A multimodal (inter)action analysis. In Kecskes Istvan (ed.), *Cambridge handbook of intercultural pragmatics*, 552–587. Cambridge: Cambridge University Press.

Norris, Sigrid & Boonyalakha Makboon. 2015. Objects, frozen actions, and identity: A multimodal (inter)action analysis. *Multimodal Communication* 4(1). 43–59.

Norris, Sigrid & Jesse Pirini. 2016. Communicating knowledge, getting attention, and negotiating disagreement via video conferencing technology: A multimodal analysis. *Journal of Organizational Knowledge Communication* 3(1). 23–48.

Payne, Scott & Paul J. Whitney. 2002. Developing L2 oral proficiency through synchronous CMC: Output, working memory, and interlanguage development. *CALICO Journal* 20(1). https://doi.org/10.1558/cj.v20i1.7-32.

Satar, Müge, Mirjam Hauck & Zeynep Bilki. 2023. Multimodal representation in virtual exchange: A social semiotic approach to critical digital literacy. *Language Learning & Technology* 27(2). 72–96.

Satar, Müge & Ciara R. Wigham. 2020. Delivering task instructions in multimodal synchronous online language teaching. *Alsic* 23. https://doi.org/10.4000/alsic.4571.

Satar, Müge & Wigham Ciara R. 2022. *Full size figures, extracts, and tables in the book titled instruction giving in online language lessons: A multimodal (inter)action analysis published by Routledge focus Applied Linguistics Series*. Newcastle upon Tyne: Newcastle University. Figure.

Satar, Müge & Wigham Ciara R. 2023. *Instruction giving in online language lessons. A multimodal (inter)action analysis*. New York and London: Routledge.

Scollon, Ron. 1998. *Mediated discourse: The nexus of practice*. London: Routledge.

Sloetjes, Han & Peter Wittenburg. 2008. Annotation by category – ELAN and ISO DCR. In *Proceedings of the 6th international conference on language resources and evaluation (LREC 2008)*. Marrakesh, Morroco: Language Resources Association. Available at: http://www.lrec-conf.org/proceedings/lrec2008/pdf/208_paper.pdf.

Stickler, Ursula & Regine Hampel. 2019. Qualitative research in online language learning: What can it do? *International Journal of Computer-Assisted Language Learning and Teaching* 9(3). 14–28.

Strauss, Anselm L. & Corbin Juliet M. 1998. *Basics of qualitative research: Techniques and procedures for developing grounded theory*, 2nd edn. Los Angeles, London, New Delhi, Singapore, Washington DC and Boston: SAGE.

Wertsch, James V. 1998. *Mind as action*. Oxford: Oxford University Press.

Wigham, Ciara R. & Müge Satar. 2021. Multimodal (inter)action analysis of task instructions in language teaching via videoconferencing: A case study. *ReCALL* 33(3). 195–213.