DE GRUYTER OPEN

Yearbook of the Poznań Linguistic Meeting 2 (2016), pp. 141–165
DOI: 10.1515/yplm-2016-0007

Concepts in multimodal discourse analysis with examples from video conferencing

Sigrid Norris
Auckland University of Technology
sigrid.norris@aut.ac.nz

Abstract

This article presents theoretical concepts and methodological tools from multimodal (inter)action analysis that allow the reader to gain new insight into the study of discourse and interaction. The data for this article comes from a video ethnographic study (with emphasis on the video data) of 17 New Zealand families (inter)acting with family members via skype or facetime across the globe. In all, 84 social actors participated in the study, ranging in age from infant to 84 years old. The analysis part of the project, with data collected between December 2014 and December 2015, is ongoing. The data presented here was collected in December 2014 and has gone through various stages of analysis, ranging from general, intermediate to micro analysis.

Using the various methodological tools and emphasising the notion of mediation, the article demonstrates how a New Zealand participant first pays focused attention to his engagement in the research project. He then performs a semantic/pragmatic means, indicating a shift in his focused attention. Here, it is demonstrated that a new focus builds up incrementally: As the participant begins to focus on the skype (inter)action with his sister and nieces, modal density increases and he establishes an emotive closeness. At this point, the technology that mediates the interaction is only a mundane aspect, taken for granted by the participants.

Keywords: human–computer interaction; language and interaction; mediation; multimodal discourse analysis; multimodal (inter)action analysis.

1. Introduction[1]

Multimodal (inter)action analysis (Norris 2004, 2009, 2011a, 2011b, 2013) is a theory of human communication with an abundance of methodological tools to

[1] This project is conducted by the Multimodal Research Centre at Auckland University of Technology, New Zealand with Sigrid Norris as PI; Jarret Geenen, Madeline Henry, Keely Kidner, Ewa Kusmierczyk, and Jesse Pirini as Researchers. Data collection and partial analysis has been co-

empirically investigate interaction. Growing out of applied linguistics, anthropological linguistics, sociolinguistics, discourse analysis, and socio-cultural psychology (Goffman 1963, 1974; Gumperz 1982; Tannen 1984; Schiffrin 1987; Hamilton 1998; Scollon 1997; 1998; 2001; van Lier 1996; Wertsch 1998; Wodak 1989) and strongly influenced by social semiotic thought (van Leeuwen 1999; Kress 2000; Kress and Van Leeuwen 1998, 2001), multimodal (inter)action analysis (Norris 2004, 2011a) is a multimodal discourse approach. Whereas some scholars in applied linguistics (Shohamy and Gorter 2008), pragmatics (Herring et al. 2013) and sociolinguistics (Bucholtz and Hall, forthcoming) are calling for or are incorporating multimodal data, this article offers a novel framework (Norris 2004, 2011) that opens up the study of discourse and interaction in vastly different ways than does the mere inclusion of multimodal data into a linguistic study.

Multimodal (inter)action analysis (Norris 2004, 2011a) differs in substantial ways from most other *discursive* approaches as well as from other *multimodal* approaches: In multimodal (inter)action analysis, language and other modes are not viewed as phenomena that exist outside of the individual to be studied as entities in and by themselves. Rather, multimodal (inter)action analysis champions to investigate language and other modes as part of the *individuals in the world* and thus, more accurately, as part of the action that the individuals perform with others, the environment, and objects within. Certainly, no one will disagree with the fact that language and other modes are part of individuals or disagree with the fact that humans are a part of their socio-cultural world acting in and with it. But linguistic theories as well as other multimodal theories fall short of explanatory tools that allow for the analysis of exactly *how* social actors, world and objects connect. Too often we read that language constructs the social at the same time as language is constructed by the social (Schiffrin 1994) and while this is certainly true, the question remains: *How do we analyse this fact in detail?*

Multimodal (inter)action analysis, based in an understanding of mediation as advocated by Wertsch (1998) and Scollon (1998, 2001), builds a framework for such detailed analysis. In this framework, every action is taken to be mediated in multiple ways.

---

funded by The Faculty of Design and Creative Technologies, The School of Communication Studies, and The Digital Mobility Research Group of the New Zealand Work Research Institute, Auckland University of Technology, New Zealand. We would like to thank all of the participants for their involvement in this research project.

> A mediated action focuses on two elements: the agent and the mediational means, emphasizing an inherent irreducible tension between the two.
>
> (Norris and Jones, 2005: 17)

All actions are thus *mediated* because social actor(s) always act with or through mediational means/cultural tools (Wertsch 1998; Scollon 1998). The notion of *mediated action* makes the concept of mediation, psychological as well as physically embodied and physically through objects and the environment, a highly important concept. Through the underlying concept of mediation in all respects of action, the framework allows for the simultaneous theoretical inclusion of social actors and their psychological make-up, objects, and the environment. The notion of mediation in this framework facilitates the resolution of differences between human actors, the things they use, and the world that they inhabit (Norris 2013). Thus, in multimodal (inter)action analysis, the notion of mediation is a theoretical concept that allows for the theoretically comprehensively bringing together of cognitive and socio-psychological, embodied physical, object physical, and environmental physical aspects *into one framework*. Through the inclusion of all of these facets, the theoretical framework embraces the complexity of interaction. In order to analyse this complexity in practical terms, various methodological tools have been developed (Norris 2004, 2009, 2011a, 2014, forthcoming; Geenen 2013; Makboon 2015; Pirini 2016), taking the study of interaction and language in use to a deeper level.

This article explicates some key concepts and methodological tools, by illustrating these through examples from a large-scale study of 17 New Zealand families (84 individuals in age from infant to 84 years old) interacting via videoconferencing technology with family members across the globe, using either skype or facetime. During the research sessions, New Zealand families used a researcher-provided laptop that recorded the online interactions. A stationary video camera positioned in the New Zealand participants' home recorded the video conferencing interactions from an external point of view, and one to three researchers (depending on availability) were present, observing the interactions and/or taking fieldnotes. The data was/is then logged according to the steps outlined in Norris (forthcoming) and is currently being analysed using multimodal (inter)action analysis (Norris 2004, 2011a), building upon general philosophical and theoretical concepts as exemplified below. Data analysis is still ongoing, but the data for this article, one of the first interactions recorded, has gone through all of these steps of analysis.

1.1. General philosophical and theoretical concepts

The usefulness of Merleau-Ponty's (1962, 1963) philosophical point of view, which states that the human being is a part of the world acting in and with it, erasing the internal/external duality is particularly evident when examining human-computer interactions. Figure 1, for example, shows a moment where Mic, a New Zealand participant in our study, (inter)acts with his environment and the objects within. Here, the left side of the image shows the larger part of Mic's computer screen, the top right illustrates a different part of his computer screen (where he will later see his own image), and the bottom right shows Mic from a video camera positioned on a tripod in his home. Mic's right hand is placed on the touchpad of the computer and his right middle finger has just pushed onto it as he is attempting to re-connect with family members in Australia.

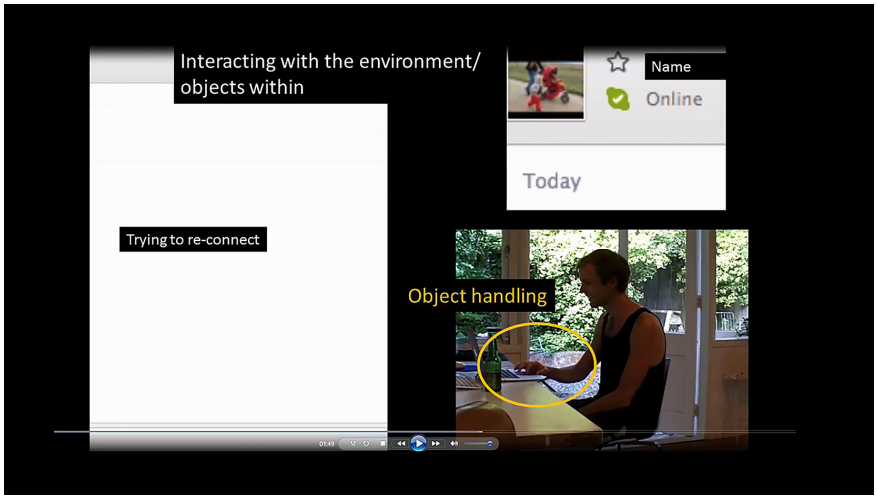

Figure 1. (Inter)acting with an object.

Without his (inter)action with the objects, the computer and touchpad, he would not be able to establish a new connection in order to then (inter)act with his sister and her children. But besides handling the object, he also (inter)acts with his environment in other important ways. Figure 2 illustrates the very next moment, when the connection is being established.

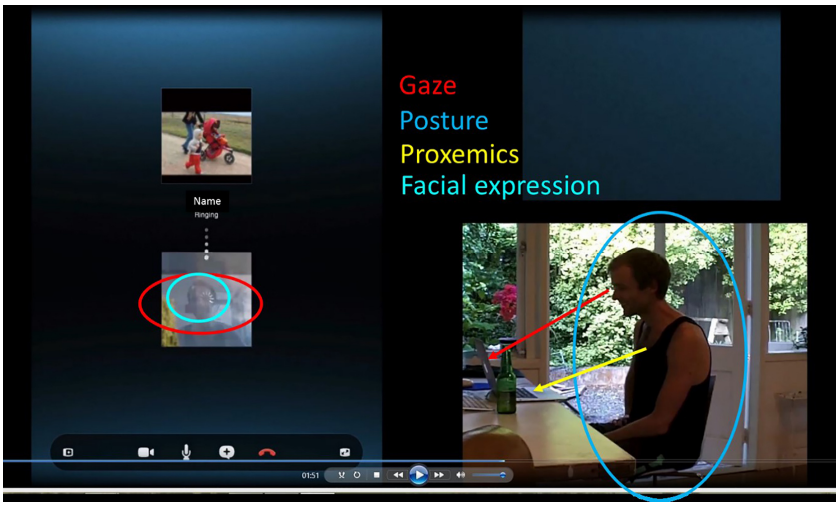

Figure 2. (Inter)acting with the environment and objects within.

In Figure 2, we see Mic gazing at the screen. Here, we observe him on the left in Figure 2 (circled in red) as he sees himself on screen and to the right, we see him sitting at the desk from the in-room camera view. He is sitting similarly as in Figure 1, but here his body shows a slightly more relaxed position with his right hand now placed on his right leg. His proxemics to the computer screen are about the same as in Figure 1, which is close enough for him to easily manipulate the computer mouse and keypad, and also far enough away to leisurely watch and be seen on screen. His facial expression that is visible on screen (left in Figure 2 circled in light green), is happy and relaxed. All of his embodied modes express his waiting and anticipation of the new connection to be established at the same time as the computer makes a ringing sound indicating the call to Australia and showing a waiting signal as droplets are moving towards the name of the call recipient, both of which Mic appears to be watching.

Soon, the receiver has taken the call (Figure 3), the ringing stops and an image appears in its place. Here, in Figure 3, we see the connection being made on the left of the image, the screenshot of the participant as he sees himself is now visible top right, and the in-room camera view of the participant is again located at the bottom right. However, the connection is not quite established yet, and we see Mic's face has changed from a full smile a moment earlier (Figure 2) to a slight worry as the connection might fail at this point.

Figure 3: A possible point of failure.

Here in Figure 3, it becomes highly evident that human beings, as Bateson (1972) pointed out, are ecologically interdependent with as well as dependent upon the environment. Only if the connection becomes established, will an (inter)action between brother and sister (or uncle, nieces and nephew) unfold. The awareness of his dependency on technology that goes beyond computer and software, which is taken for-granted and is largely ubiquitous as soon as a working connection is established, is here in Figure 3 present and visible in the participant's facial expression.

These examples illustrate the notion of social actors as part of the world, acting in and with it. In multimodal (inter)action analysis social actors are, other than in actor network theory (Latour 2005), always and only humans. The computer in the above examples is a cultural tool/mediational means, and so are the software and the many hidden technologies that make the connection between social actors possible.

Mediation is a term that is often used in regards to technology as computer mediated communication regards any kind of communication mediated by one or more technological devices. In multimodal (inter)action analysis however, technology in the above example is only one aspect of mediation: For example, as Mic operates the touchpad (Figure 1), he utilises the cultural tools (laptop,

skype, broadband connection, and other ubiquitous technologies) in order to connect to his family in Australia. Clearly, this is the kind of mediation that many researchers have in mind when speaking of computer mediated interaction. However, as we will see below, multimodal (inter)action analysis with its roots in mediated discourse analysis, takes mediation as theoretically much more important than other frameworks.

2. Multimodal (inter)action analysis: An interdisciplinary approach

Multimodal (inter)action analysis (Norris 2004, 2009, 2011a, 2011b, 2013, 2014, 2015) originating from mediated discourse analysis (Scollon 1998, 2001) is based in the sociological interest of humans acting in the world that we find in the work of Goffman (1963); incorporates the interest in intercultural interaction that we find in the work of Gumperz (1982); includes an interest in power in interaction that we find in the work of Wodak (1989); delves into the micro-analysis of interaction that we find in the work of Tannen (1984), Schiffrin (1987), or Hamilton (1998); has a strong interest in applied linguistics that we find in the work of van Lier (1996); is strongly influenced by socio-cultural psychology as we find in the work of Wertsch (1998); and is grounded in social semiotic thought that we find in the writings of van Leeuwen and Kress (van Leeuwen 1999; Kress 2000; Kress and van Leeuwen 1998, 2001). With these foundations, multimodal (inter)action analysis (Norris 2004, 2011) has developed into a strong theoretical framework with an abundance of methodological tools (Norris 2004, 2009, 2011, 2013a, 2013b, 2014, forthcoming; Geenen 2013; Makboon 2015; Pirini 2015, 2016) that make the analysis of (always) multimodal (inter)action possible, opening up research into new and promising directions.

As mentioned above, a main theoretical notion in this framework is the concept of mediation. The importance of mediation finds itself in the unit of analysis, the mediated action, which has been adopted from Wertsch (1998) (who developed it from Vygotsky) and Scollon (1998) (who developed it from Wertsch), and is further developed and thereby delineated into three methodological tools by Norris (2004). Theoretically, the mediated action is defined as social actor(s) acting with or through cultural tools/mediational means (Wertsch 1998; Scollon 1998). The mediated action as unit of analysis incorporates the social actor(s) *and* the (always multiple) cultural tools/mediational means. Thus human(s) +

cultural tools with their always present inherent tension *build* the unit of analysis. The terms cultural tools and mediational means are used interchangeably as mediational means are cultural and cultural tools mediate action. This theoretical concept of mediation is embraced in the conception of the three methodological units of analysis, the lower-level mediated action, the higher-level mediated action, and the frozen mediated action.

2.1. The concepts *lower-level*, *higher-level* and *frozen mediated actions*: Units of analysis

Multimodal (inter)action analysis conceives of all actions as mediated actions. Therefore, as soon as we speak of lower-level, higher-level, or frozen actions, we speak of mediated actions (even if it is not always stated explicitly). The lower-level mediated action is defined as the smallest pragmatic meaning unit of a mode (Norris, 2004). For example, an utterance is the smallest meaning unit of the mode of spoken language. An utterance is a lower-level mediated action as it is produced by a social actor + multiple socio-cultural and psychological, embodied and physical, and semiotic mediational means/cultural tools as an utterance is mediated by mediational means/cultural tools such as the larynx, lips, teeth, tongue, out-breath, a language system, knowledge, and socio-cultural relevance. By theorizing that every lower-level action, no matter what it entails, is mediated in multiple ways, we can see that computer mediation in human-computer interaction is not so very different from the mediation involved in the production of an utterance. Revisiting Figure 1, where Mic pushes the touchpad, this lower-level action (or smallest pragmatic meaning unit of the mode of computer use) is also mediated by multiple socio-cultural and psychological, embodied, physical, and semiotic mediational means/cultural tools. Here, the action of pressing onto the touchpad is mediated by the finger, the hand/arm/ body posture (to allow for the finger movement), the laptop and its touchpad, the ubiquitous technological tools effecting a change through this finger movement, the knowledge about the device and the result of this action, and so on. While in practical terms, the mediation in the production of an utterance is vastly different from the mediation involved in the pushing onto a touchpad, theoretically speaking, we clearly can see that there exist great similarities as well; as each lower-level action performed by a social actor is mediated by multiple socio-cultural and psychological, embodied and physical, and semiotic mediational means/cultural tools.

In line with this, the term *mode* in multimodal (inter)action analysis, is defined as a *system of mediated action* (Norris 2013), incorporating a psychological, physical, socio-cultural and with it a historical dimension to the concept and adhering to the theoretical notion of mediated action. Conceived of as systems of mediated action (Norris 2013), modes are learned by social actors in and through contact with other social actors, the environment and objects within. In this definition, the complexity of modal use in interaction is embraced at the very same time as the always multiple mediation and the inherent tension between social actor(s), environment and objects within are contained.

Lower-level mediated actions are methodological tools that allow researchers to delineate micro actions that are (almost) never delineated by social actors in their everyday lives. We may, of course, find the deliberate performance of a blinking of the eyes or a loud outbreath or the push of a touchpad, but such instances of individual lower-level actions are still always performed together with other lower-level actions, some in *and* some out of synchrony, within the performance of higher-level actions. For example, the lower-level action of pushing the touchpad in Figure 1 is performed intentionally, but this action is performed together with other lower-level actions such as a smile and gaze as illustrated on Figure 4 (Figure 1 revisited).

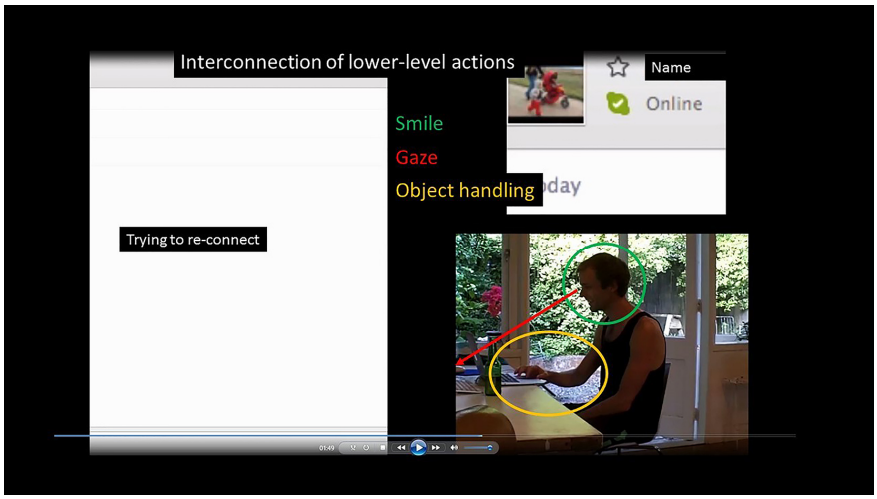

Figure 4. Interconnection of lower-level actions.

Higher-level mediated actions are those actions that social actors usually intend to perform and/or, as explained in more detail below, are aware of and/or pay attention to. Higher-level actions come about through the coming together of many chains of lower-level actions (such as several utterances chained together by speakers, gaze shifts, postural shifts and so on) at the same time as the higher-level actions constitute these lower-level actions. Thus, lower-level and higher-level mediated actions always constitute each other. Figure 5 illustrates this point.

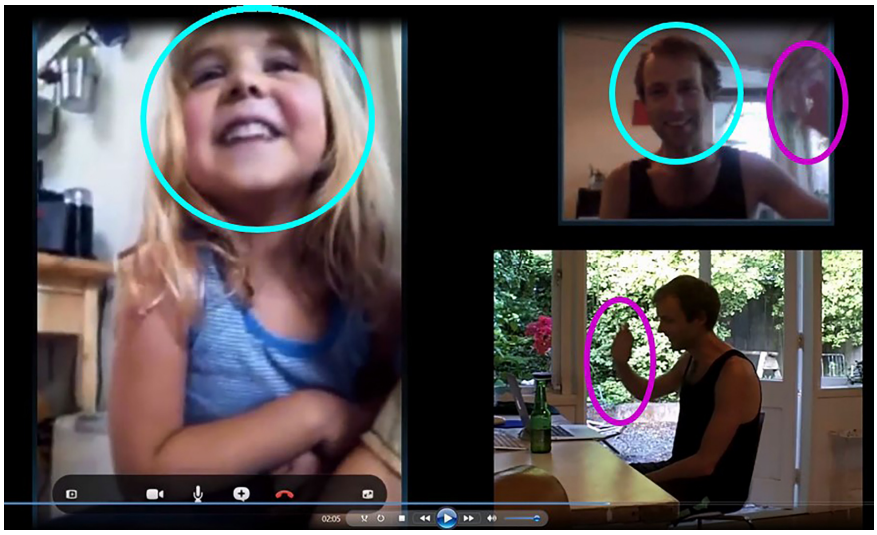

Figure 5. Lower- and higher-level mediated actions constitute each other.

As we see in Figure 5, the connection has been established and the uncle's worried expression from just a moment earlier turns into a smile at the same time as he begins to wave to his niece in Australia and the niece in Australia simultaneously smiles at her uncle. All of these lower-level mediated actions, each one of which is mediated in multiple ways, are part of the higher-level mediated action of these participants interacting via skype. Here, it becomes apparent that mediation of this higher-level action, the skype interaction, is anything but simple. Rather, we find that a higher-level action such as this skype interaction is medi-

ated in vastly complex ways. While in much research on computer mediated communication, technological mediation is discussed or referred to, multimodal (inter)action analysis demonstrates that mediation on the one hand goes *far beyond* technological mediation, opening up the study of technology-mediated interaction in new directions; and on the other hand, illustrates that technology-mediated interaction is theoretically not all that different from other kinds of interaction because all interaction is complexly mediated, opening up the study of interaction in new directions.

In multimodal (inter)action analysis, we can dissect a higher-level action and the multitude of mediation; or we can dissect a higher-level action and illustrate how it is made up and simultaneously produces a multitude of chained lower-level mediated actions, that a social actor may or may not be focused upon. The more focused upon a higher-level action a social actor is, the stronger is the higher-level action's modal make-up. Strength of a higher-level action's modal make-up is represented through the concept of modal density which is discussed in the next section.

But, briefly revisiting Figure 5, it is important to note that neither the wave nor the smiles or the evolving utterances are separated from each other by the participants in interaction; it is exactly their coming together that makes this video-conferencing session just that: a video conferencing-session. Besides the lower-level and the higher-level mediated actions, the third unit of analysis is the frozen mediated action in multimodal (inter)action analysis. This concept allows for the analysis of relevant actions that have been performed by a social actor at an earlier time, which become frozen in objects or the environment. As a quick example, when we have a look at Figure 5 once more, we see a beer bottle standing on the desk (in the lower right image of the screen grab). This bottle tells of the action of Mic drinking a beer and having positioned it where it is standing now. Even if we had not witnessed him at points in the video having a sip of beer now and again, we would read the action of him drinking beer off of the object itself. As discussed elsewhere (Norris 2004), usually social actors read those actions off of objects that are closest in time and space to the object and the individual. These read-off actions may or may not be correct and are in interaction often confirmed or rejected and corrected. As we will see in section 2.4 below, the concept of frozen action, just as the concept of lower-level action and the concept of higher-level action, is highly relevant when analysing interaction.

2.2. The concepts *modal density* and *foreground-background levels of attention*

Modal density = lower-level action density within a higher-level action (Norris 2004, 2008, 2009, 2011). The concept of modal density allows to analyse interactions beyond the focus; and the concept of a *foreground-background continuum* allows to visually represent the various levels of attention that an individual is simultaneously engaged in. Revisiting the example given in Figure 3, more information is necessary to allow for the analysis of Mic's attention levels at that very moment as shown in Figure 6.

Figure 6 illustrates that Mic is engaged in three simultaneous higher-level actions: (1) He is skyping with family members in Australia; (2) He is engaged in a research project; and (3) He is interacting with his girlfriend. The first higher-level action, the moment of reconnecting with his sister and nieces in Australia has briefly been discussed above (Figure 3). Mic's skype call, as mentioned in the Introduction, is part of a research session, in which Mic is using a research laptop that records his online interaction, an external camera that records him from an in-room point of view, and two researchers, who are observing him from the back of the room. Simultaneously, and from before the time when the researchers arrived at his house, Mic's girlfriend is present. Mic, no doubt is aware of all of this as he is sitting in front of the laptop trying to reconnect with his sister in Australia. However, Mic is not aware of or paying attention to all of these higher-level actions to the same degree. Here, as Mic is waiting for the connection to be established, he is highly aware of the research session. When looking at Figure 6, we see that at this very moment the research session modally dominates: Mic takes up close proxemics to the research laptop and he is well aware of being recorded; he is aware of his proxemics to the stationary camera and of the fact that this camera too records him; and he is aware of the presence of the researchers due to his proxemics to them and having spoken with them just a moment before. Taking part in a research project and the many mediated actions that this entails (which are now frozen in the objects: laptop, tripod, camera, researchers' notebooks, etc.) as well as Mic's embodied modes of posture and his bodily proxemics to the objects that entail the frozen actions and to the researchers present in the room, cumulate in high modal density as illustrated in Figure 6, demonstrating that he is focusing on his engagement in the research session at this moment. At the same time, and as mentioned previously, Mic is paying attention to skype as he is waiting for the connection to be made. His lower-level actions of a worried facial expression, direct gaze at the com-

puter screen, posture (positioned to easily see and be seen), relaxed arms/hands all cumulate in medium modal density as illustrated in Figure 6, demonstrating that he is engaged in the skype call in the mid-ground of his attention. Still simultaneously, but to a much lesser degree, Mic is aware of the presence of his girlfriend and his interaction with her. For example, he turns to her later and requests her to join him in his skype interaction. However, at this very moment, it is her proxemics to him and her presence in the room that cumulate in a low modal density as illustrated in Figure 6, demonstrating that Mic is paying least attention to the interaction with her at this time.

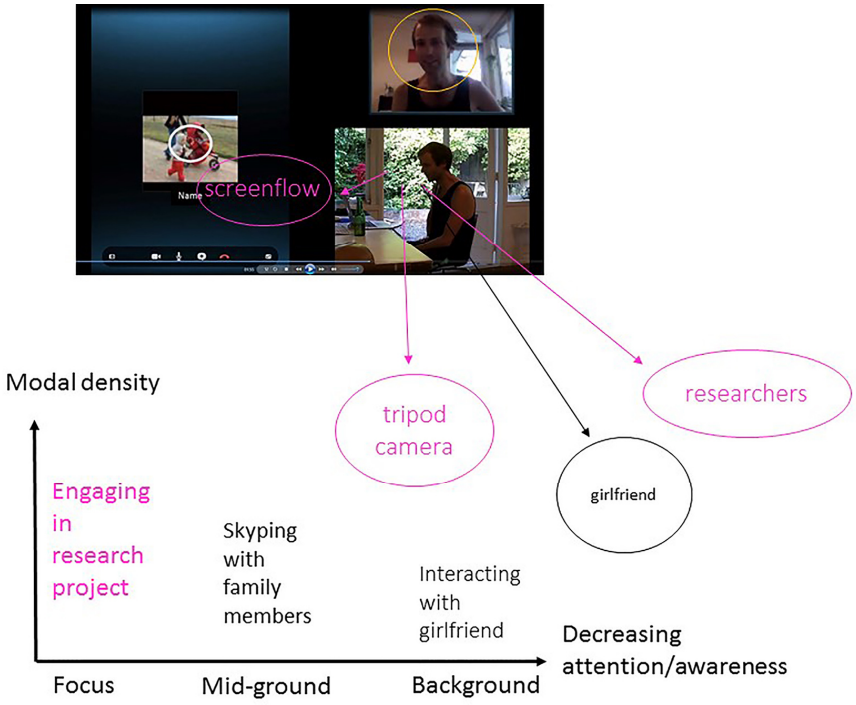

Figure 6. The various interactions that Mic is engaged in at a particular point in time.

Mic's focused attention/awareness of taking part in the research project persists for some time. But at almost 4 minutes into the skype session, Mic indicates a

change in focus, which is analysable through the concept of semantic/pragmatic means outlined in the next section. As he refocuses, Mic becomes more engaged in the skype interaction as the modal density of this higher-level action rises.

2.3. Concept semantic/pragmatic means

Semantic/pragmatic means are pronounced lower-level actions that indicate a change in focus by the one producing them (Norris 2004, 2011). These means are semantic in that they produce a change in meaning of the higher-level actions in the attention levels of the performer; and they are pragmatic as their use produces a knowledge of that change in attention to a different higher-level action for others engaged in interaction. Semantic/pragmatic means are always pronounced and have a structuring function. As such, they sit somewhat *outside* of the higher-level actions themselves. When Mic produces the semantic/pragmatic means of bowing forward (Figure 7), he indicates a shift from paying focused attention to his engagement in the research project to paying focused attention to the interaction with his sister and nieces in Australia. Here, *bowing down low* (a pronounced lower-level action) does not convey meaning as a part of the higher-level action of engaging in the research project, nor does it convey meaning that connects to the higher-level action of interacting with his sister and nieces via skype.

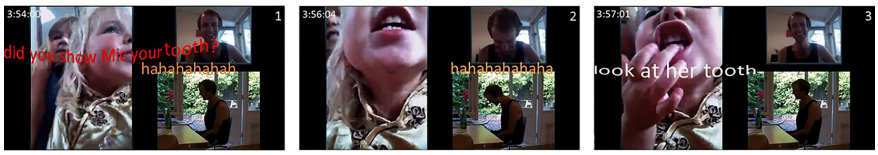

Figure 7. Semantic/pragmatic means: Bowing forward.

Social actors, who are engaged in multiple higher-level actions, quite frequently shift their focused attention from one to another higher-level action that they are involved in. Refocusing is always structured by semantic means, as the social actor is restructuring not only the attention that they are paying but also the meaning that they are constructing by focusing on a particular higher-level action. As the means that structure attention and meaning in the mind of the social

actor producing them is always visible or audible, these means also function pragmatically in interaction so that others are often aware of what someone else is focusing on (Norris 2004, 2006, 2011a).

As is visible in the brief transcript in Figure 7 image 1 (reproduced larger in Figure 8), Mic's sister is prompting 3-year old Sophie indirectly to show Mic her tooth when she says *did you show Mic your tooth?* (see Geenen, forthcoming for a detailed analysis of Sophie's (inter)action). Mic, however, is still laughing at something that occurred earlier in the skype conversation, and he is still focused upon the research session. However, as he continues to laugh, he now bows his head low (Figure 7 image 2) in a semantic/pragmatic means, and when his 5-year old niece Isla directs him to *look at her tooth* (to look at Sophie's tooth) Mic's facial expression changes and illustrates that he is now focused upon the skype interaction with his sister and nieces in Australia as visible in the transcript (Figure 9) discussed in the next section.

3. How do these concepts work together?: *A shift in focus*

As discussed in Section 2.2, Mic is first focused upon the research session, he mid-grounds the skype interaction, and backgrounds the interaction with his girlfriend (Figure 6). This analysis was conducted through the concepts of lower-level, higher-level and frozen mediated actions, modal density, and the foreground-background continuum of attention/awareness. Utilizing the concept of semantic/pragmatic means, it was then illustrated in Section 2.3 that it is possible to delineate the exact point at which Mic changes his focus from being engaged in a research project to interacting via skype with his sister and nieces due to the analysis of a semantic/pragmatic means (Figure 7). In Figure 9 below, Mic's new focus becomes apparent as we again utilise the concepts of lower-level, higher-level and frozen mediated actions as well as modal density and the foreground-background continuum of attention/awareness.

The multimodal transcript (Figure 9) follows the transcription conventions described in Norris (2002, 2004, 2011) with a reading path from left to right and top to bottom. Each individual screengrab is numbered top right and the exact time in the video recording is presented top left of each screen grab; utterances by individual participants are colour coded, overlaid on top of the screengrabs to illustrate the coming together of spoken language and other modes and highlighting the rising and falling of intonation as produced by the speaker as illustrated in Ladefoged (1975). In the following transcript, we see Mic's sister's

(the girls' mother's) utterances in red. She is not visible in the images. Then, we find 5-year old Isla's utterances in white and she is only visible in the first and last two images of the transcript, but her hand is clearly visible in images 7–10. Sophie is visible in all screen grabs but she does not speak in this excerpt; and Mic is clearly visible and his utterances are produced in yellow as shown in Figure 8.

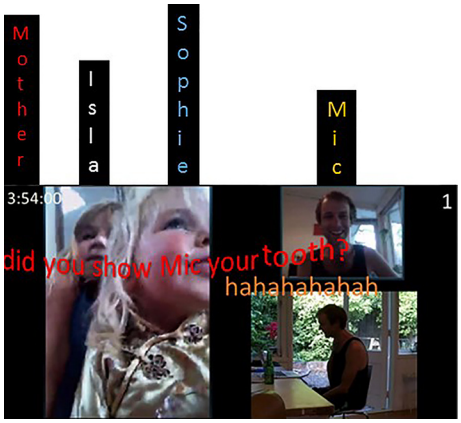

Figure 8. Social actors and their colour-coded utterances in the transcript in Figure 9: Mother's utterances in red; uncle's (Mic's) utterances in yellow; and Isla's utterances in white.

The first three images in Figure 9 repeat the images in Figure 7 as they illustrate on the one hand that a new topic is broached by the Sophie's mother (Mic's sister) and that Mic is not immediately responding to this topic as he is still focused upon the research session. As he refocuses, Mic becomes visibly more engaged in the skype interaction demonstrating that modal density of this higher-level action rises.

In the first three images in Figure 9, Mic performs his sematic/pragmatic means and in image 4 we see how modal density begins to rise. Social actors often lag once they have performed a semantic/pragmatic means (Norris 2004, 2011a) before they are fully engaged in the newly focused upon higher-level action. What this shows is that social actors often take some time before they build up the modal density and when examining these changes in great detail, we can

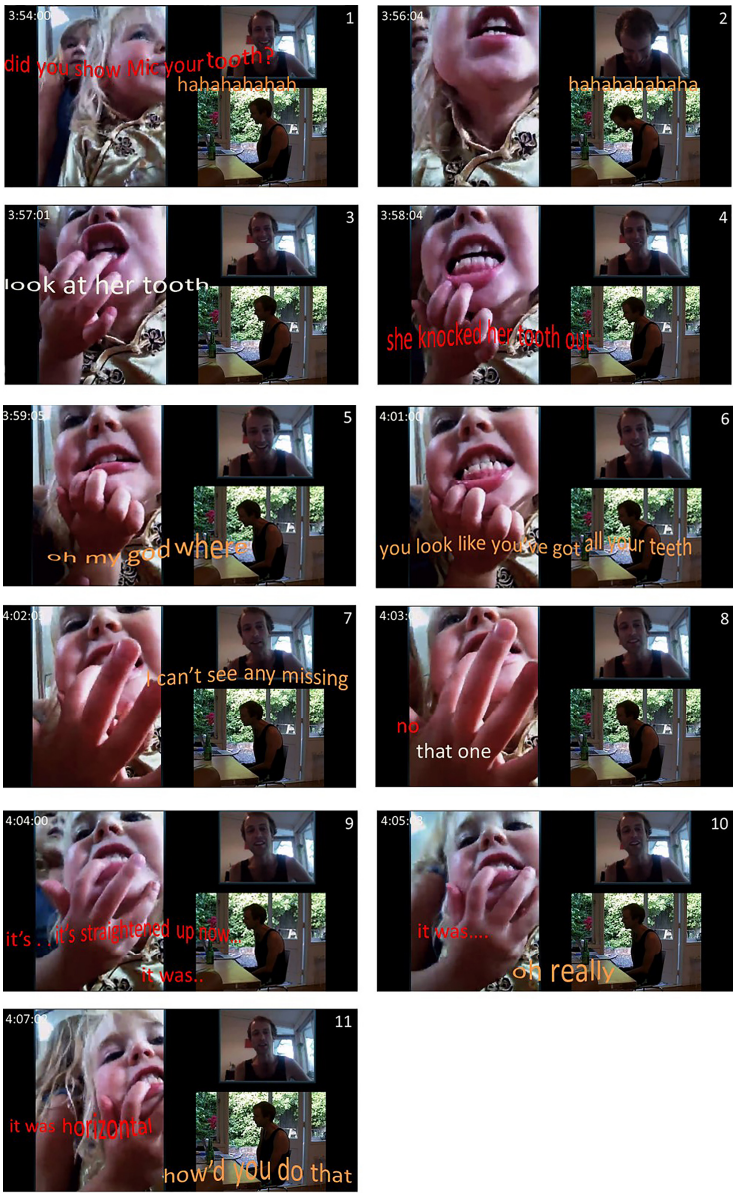

Figure 9. Mic is now fully focused upon the skype interaction (images 5–10). This same excerpt is analysed in Geenen (forthcoming), detailing Sophie's learning of making a relevant interactive contribution in family interaction.

see how modal density is built up incrementally. In image 4 of Figure 9, only a little over a second after the indicated shift in focus, we see a small change in Mic's facial expression and head movement: his previous smile turns into a serious expression and his head moves forward and down a bit. Then, in image 5, another second later, Mic has moved his posture and with it his head further forward, is now gazing intently at Sophie's teeth displayed on his screen, and speaks, ending quite loudly, saying *oh my God where*. As Sophie pulls down her lower lip, Mic continues to look intently, beginning to tilt his head and saying *you look like you've got all your teeth* (image 6). However, his facial expression displays that he is unsure as he tilts his head further and continues to intently gaze at the teeth and Isla's hand makes her way to Sophie's tooth (image 7). In image 8, Isla is pointing at a specific tooth in Sophie's mouth; her mother says *no*, and Isla latches this *no* of her mother saying *that one*. As they are producing the utterances and Isla is pointing, we can see in Mic's facial expression the pain that he is feeling by the mere thought of Sophie having knocked out a tooth. Mic moves his head and posture back a little as if to move away from a blow; his head is still tilted and the facial expression is expressing even more pain now with his mouth showing his teeth, the edges of his lips pulled downward, and his eyes squinted (image 9). Mic continues to move back slightly and continues to produce the facial expression when his sister says *it was* (image 10) and he exclaims *oh really* and she continues with *it was horizontal* (image 11); and Mic questions Sophie *how'd you do that*.

By analysing the lower-level actions produced, we can demonstrate that Mic's change in focused higher-level action comes about after the production of a semantic/pragmatic means which re-structures the amount of attention that he pays to the simultaneous higher-level actions that he is engaged in *and* which indicates this restructuring to others. However, rather than occurring immediately, a shift is produced incrementally (see also Norris 2008) with modal density building up through (inter)action. In the above example, Mic begins building up modal density of the higher-level action of (inter)acting with his sister and nieces via skype through embodied lower-level actions (and chains thereof) such as facial expression, head movement, postural change, gaze, and proxemics to the laptop screen as well as the production of utterances. As modal density of this higher-level action increases, modal density of the higher-level action of engaging in a research project decreases in tandem. Resultantly, Mic's attention levels can now be visualised with the concept of the modal density foreground-background continuum (Figure 10).

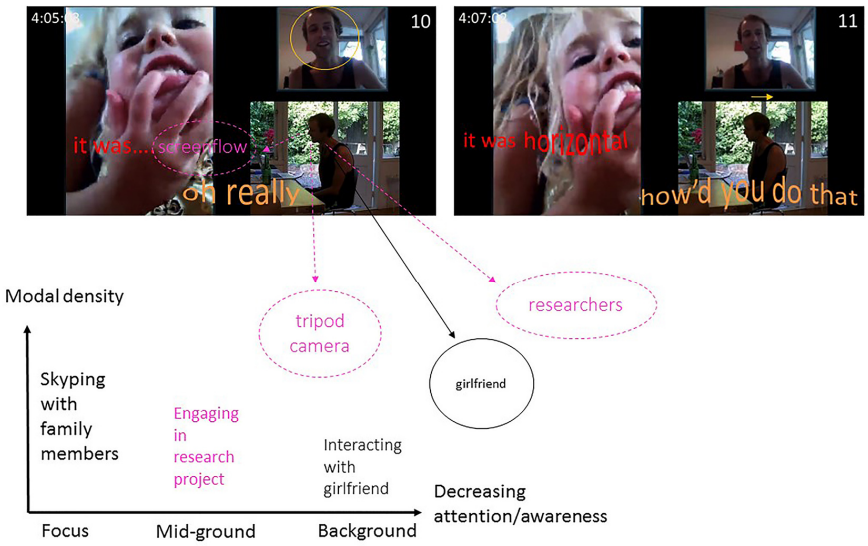

Figure 10. New distribution of higher-level actions in Mic's attention levels.

The graph in Figure 10 visualises the new distribution of attention/awareness of the higher-level actions that he is involved in. As illustrated in Figure 9 previously, Mic begins to focus more and more on the higher-level action of skyping with family members as he is concurrently paying less attention to the higher-level action of engaging in a research project. As lower-level mediated action density to produce the higher-level action of skyping with family increases, the lower-level action density for the higher-level action of engaging in the research project decreases. With a shift in focus, Mic's gazing at the laptop screen is related more to the looking at the damaged tooth and his awareness of being recorded diminishes. Correspondingly, with modal density increasing to produce the higher-level action of skyping with family members through the many embodied modes that Mic uses, the modal density produced by the frozen actions embedded in the tripod and camera as well as the physical presence and proxemics to the researchers diminish with his paying less attention to them (indicated by dotted lines in Figure 10), thereby pushing the higher-level action of engaging in a research project to the mid-ground of Mic's attention/awareness.

The modal density foreground-background continuum, although a two-dimensional and relatively simplistic visualisation, allows us to *map* the very

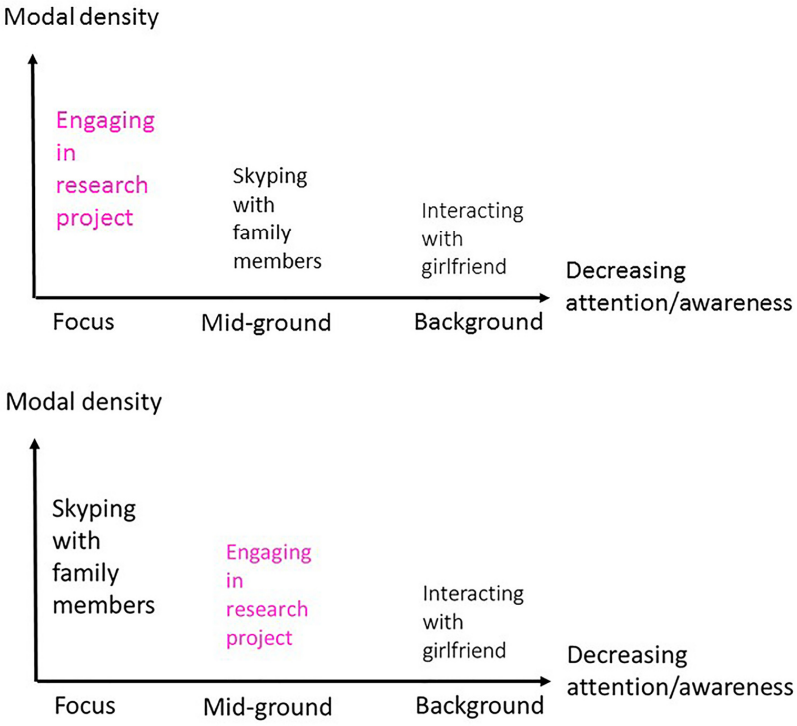

Figure 11. Modal density before and after Mic's performance of the semantic/pragmatic means.

complexly performed change in Mic's attention/awareness in order to clearly demonstrate the shift that has taken place (Figure 11).

4. How do the concepts work together?: *Mediation*

In this section, the above example (Figure 9) is revisited with an emphasis on mediation: Each lower-level action performed by a social actor is mediated by multiple socio-cultural, cognitive and psychological, embodied, physical, and semiotic mediational means/cultural tools. The sematic/pragmatic means that Mic performs in the first three images of Figure 9 is mediated psychologically

as he appears to feel more comfortable to change his focus away from the research session onto the actual skype interaction; it is mediated cognitively, as the means itself indicates a cognitive re-structuring of Mic's focused attention; the semantic/pragmatic means (the bowing of his head) is mediated socio-culturally as it is learned through social and cultural development; the means is mediated by his physical body, particularly his head; and it is mediated semiotically as the bowing of the head at this moment in interaction is meaningfully produced as a structuring device and can be read by others as a shift in his focus.

In image 4 of Figure 9, where Mic's previous smile turns into a serious expression and his head moves forward and down a bit, he reacts to the utterance and the serious tone of his sister's voice when asking Sophie *did you show Mic your tooth?* (image 1) and then explaining to Mic *'she knocked her tooth out* (image 4). This producing of a serious expression is again mediated in multiple ways from cognitive/psychological as he realises that his sister is sharing a serious matter; it is mediated socio-culturally as a serious matter and tone of voice by one social actor in interaction is to be responded to in a serious way by the other; it is mediated embodied physically as he changes the tension in his facial muscles; and it is mediated semiotically as the facial expression displays his knowledge of these semiotic systems.

Similarly, one can work through each of the lower-level actions that Mic performs and establish the multiple ways that they are mediated. However, an intensity of modal density is also developed by the interplay of several lower-level actions and their mediation. In image 5 of Figure 9 for example, Mic continues to move forward and he gazes intently at Sophie's teeth as he says *oh my God where*, emphasizing the *where* with intensity of voice. These lower-level actions not only are each mediated in multiple ways, they also mediate each other: Mic's embodied physical postural shift forward mediates his intent gaze at Sophie's teeth; Mic's newly established closeness to Sophie's teeth and his intent gaze in turn mediate his emphasising the word *where*. As all of these lower-level actions come together, they demonstrate Mic's focus.

Then, even though Mic says *you look like you've got all your teeth* (image 6) and continues with *I can't see any missing* (image 7) in a re-assuring tone of voice, Mic's facial expression, proxemics to the screen and intensity of gaze suggest worry. Here, we see dual socio-cultural mediation of an intertwined multimodal moment, linking reassurance with worry in embodied complex ways. The physical embodied mediation allows for a skilful realisation of semi-

otic dual expression of contradictory meaning, whereby the semiotic systems of course also mediate the interactive moment.

As Mic's sister produces her *no*, and then explains that the tooth was horizontal (images 8-11) Mic's facial expression mediates his empathy, the pain that he is feeling for Sophie having damaged her tooth. His empathy is further mediated as Mic moves his head and posture back and he squints his eyes in apparent pain. Of course, each of these lower-level actions is not only mediated psychologically by his feeling of empathy, but are also mediated in embodied physical, socio-cultural, and semiotic ways.

During this time of high modal density and complex cognitive, psychological, socio-cultural, and semiotic mediation of the interaction with his sister and nieces, the computer technological mediation, which was apparent in Mic's earlier facial expression (Figure 3) is here taken for-granted and ubiquitous.

5. Conclusion

This article has explicated some key concepts of multimodal (inter)action analysis (Norris 2004, 2011a, 2015, forthcoming) using examples from a family video conferencing interaction. Multimodal (inter)action analysis is a framework with strong theoretical foundations (Wertsch 1998; Scollon 1998; 2001) and theoretically linked methodological tools that situate human social actors with their cognitive, psychological, and bodily physical dimension as always linked to their physical and socio-cultural environment. Taking the mediated action as its unit of analysis, the framework embraces the complexity and constant inherent tensions that exist in the unit of social actor(s) *plus* mediational means/cultural tools. Through this unit of analysis, and more so through the methodological tools derived from it (the lower-level, higher-level, and frozen mediated actions) the framework allows for an inclusion of all of the various multimodal dimensions. Thus it becomes possible to incorporate all modes into a discourse study; analyse the interaction as linked to the relevant settings and objects within; and to analyse the (almost) always multiple actions that social actors engage in on various levels of their attention. After having explicated some of the key concepts of this framework in the first sections, the article turned to the analysis of a brief family interaction via skype in which it was first shown that that the New Zealand participant Mic payed more attention to his engagement in the research project than to the unfolding skype interaction. This analysis is only possible because of the multiplicity of data collected: the online

recording, the stationary camera recording, and the observations made by the researchers. Such an analysis, for example, would not be possible for any of the overseas participants because for all overseas participants we only have the online data.

Next, the article showed Mic's semantic/pragmatic means (his bowing his head), which indicated a shift in focused attention. A close analysis of the emerging interaction then illustrated how Mic's modal density for the higher-level action of interacting with his sister and nieces via skype incrementally increased as the modal density for the higher-level action of being engaged in a research project decreased. Through increasing multimodal interactional complexity mediated in multifaceted ways, Mic increased modal density of the interaction with his sister and nieces and established an emotive closeness. At this time, the sharing about Sophie's damaged tooth and Mic's displayed empathy takes on great importance, while the technology that mediates the interaction is only a mundane aspect, which is taken for granted by the participants. Whereas the technology is not taken for granted at a possible point of failure (Figure 3), it here becomes ubiquitous as it functions correctly.

References

Bateson, G. 1972. *Steps to an ecology of mind*. New York: Ballantine.

Bucholtz, M. and K. Hall. Forthcoming. "Embodied sociolinguistics". In: Coupland, N. (ed.), *Sociolinguistics: Theoretical debates*. Cambridge: Cambridge University Press.

Geenen, J. 2013. Kitesurfing: Action, (inter)action and mediation. (PhD thesis, Auckland University of Technology.)

Geenen, J. Forthcoming. *Multimodal acquisition of interactive aptitudes: A microgenetic case study*.

Goffman, E. 1963. *Behaviour in public places*. New York: Free Press.

Goffman, E. 1974. *Frame analysis*. New York: Harper.

Gumperz, J. 1982. *Discourse strategies*. Cambridge: Cambridge University Press.

Hamilton, H. 1998. "Reported speech in survivor identity in on-line bone marrow transplantation narratives". *Journal of Sociolinguistics* 2(1). 53–67.

Herring, S., D. Stein and T. Virtanen (eds). 2013. *Pragmatics of computer-mediated communication*. Berlin, New York: de Gruyter Mouton.

Kress, G. 2000. "Design and transformation: New theories of meaning". In: Cope, B. and M. Kalantzis (eds.), *Multiliteracies: Literacy learning and the design of social futures*. Psychology Press. 153–203.

Kress, G. and T. Van Leeuwen. 1998. *Reading images: A grammar of visual design*. London: Routledge.

Kress, G. and T. Van Leeuwen. 2001. *Multimodal discourse: The modes and media of contemporary communication.* London: Edward Arnold.
Ladefoged, P. 1975. *A course in phonetics*. Orlando: Harcourt Brace.
Latour, B. 2005. *reassembling the social: an introduction to actor-network-theory.* Oxford: Oxford University Press.
Makboon, B. 2015. Spiritual vegetarianism: Identity in everyday life of Thai non-traditional religious cult members. (PhD thesis, Auckland University of Technology.)
Norris, S. 2004. *Analyzing multimodal interaction: A methodological framework*. London: Routledge.
Norris, S. 2006. "Multiparty interaction: A multimodal perspective on relevance". *Discourse Studies* 8(3). 401–421.
Norris, S. 2008. "Personal identity construction: A multimodal perspective". In: Bhatia, V., J. Flowerdew and R.H. Jones (eds), *New directions in discourse*. London: Routledge. 132–149.
Norris, S. 2009. "Modal density and modal configurations: Multimodal actions". In: Jewit, C. (ed.), *Routledge handbook for multimodal discourse analysis.* London: Routledge.
Norris, S. 2011a. *Identity in (inter)action: Introducing multimodal (inter)action analysis*. Berlin: de Gruyter Mouton.
Norris, S. 2011b. "Three hierarchical positions of deictic gesture in relation to spoken language: A multimodal interaction analysis". *Visual Communication* 10(2). 1–19.
Norris, S. 2013a. "What is a mode? Smell, olfactory perception, and the notion of mode in multimodal mediated theory". *Multimodal Communication* 2(2). 155–169.
Norris, S. 2013 b. "Multimodal (inter)action analysis: An integrative methodology". In: Müller, C., E. Fricke, A. Cienki and D. McNeill (eds.), *Body – language – communication*. Berlin/New York: de Gruyter Mouton.
Norris, S. 2014. "The impact of literacy based schooling on learning a creative practice: Modal configurations, practices and discourses". *Multimodal Communication* 3(2). 181–196.
Norris, S. (ed). 2015. *Multimodality: Critical concepts in linguistics*. (Vols. I–IV.) Abingdon: Routledge.
Norris, S. Forthcoming. *Working with multimodal data: Research methods in multimodal discourse analysis*. Hoboken, NJ: John Wiley and Sons.
Norris, S. and R.H. Jones. 2005. "Introducing mediated action". In: Norris, S. and R.H. Jones (eds.), *Discourse in action: Introducing Mediated Discourse Analysis.* London, New York: Routledge. 17–19.
Pietikäinen, S., P. Lane, H. Salo and S. Laihiala-Kankainen. 2011. "Frozen actions in the Arctic linguistic landscape: A nexus analysis of language processes in visual space". *International Journal of Multilingualism* 8(4). 277–298.
Pirini, J. 2015. Research into tutoring: Exploring agency and intersubjectivity. (PhD thesis, Auckland University of Technology.)
Pirini, J. 2016. "Intersubjectivity and materiality: A multimodal perspective". *Multimodal Communication* 5(1). 1–14.
Schiffrin, D. 1987. *Discourse markers.* Cambridge: Cambridge University Press.
Schiffrin, D. 1994. *Approaches to discourse.* Oxford: Blackwell.

Scollon, R. 1997. "Handbills, tissues, and condoms: A site of engagement for the construction of identity in public discourse". *Journal of Sociolinguistics* 1(1). 39–61.
Scollon, R. 1998. *Mediated discourse as social interaction*. London: Longman.
Scollon, R. 2001. *Mediated discourse: The nexus of practice.* London: Routledge.
Shohamy, E. and D. Gorter (eds.). 2008. *Linguistic landscape: Expanding the scenery.* London: Routledge.
Tannen, D. 1984. *Conversational style: Analyzing talk among friends.* Norwood, NJ: Ablex.
van Leeuwen, T. 1999. *Speech, music, sound.* London: Macmillan.
Van Lier, L. 1996. *Interaction in the language curriculum: Awareness, autonomy and authenticity*. London: Longman.
Wertsch, J. 1998. *Voices of the mind: A sociocultural approach to mediated action.* Cambridge, MA: Harvard University Press.
Wodak, R. 1989. *Language, power and ideology: Studies in political discourse*. Amsterdam, Philadelphia: John Benjamins.

Address for correspondence:
Sigrid Norris
Multimodal Research Centre
School of Communication Studies
Auckland University of Technology
Private Bag 92006
Auckland 1142
New Zealand
sigrid.norris@aut.ac.nz