

Focused attention in focus: Crossing Micro-Analytical Boundaries¹

SIGRID NORRIS

I. Introduction

This article sets out to discuss an issue that discourse analysts examining everyday interactions often do not worry about: Focused attention. The article hones in on the theoretical issue of focused attention and exemplifies it with an empirical example from a study of 82 participants interacting with family members via Skype. The issue presented in this article is two-fold:

1. The article argues that we cannot determine focused attention when we cut our data pieces too small; and
2. The article argues that focused attention cannot be determined on the sole ground that a participant is using language to communicate.

Figure 1 shows a participant in New Zealand Skyping with his sister and niece in Australia. The connection has just been established less than 15 seconds earlier when the mother of the child and the participant's (the man sitting in front of the laptop) sister prompts her ›you got to look at the screen‹ and once the child looks at the screen, the participant reacts to seeing the child's face with ›there you are‹.

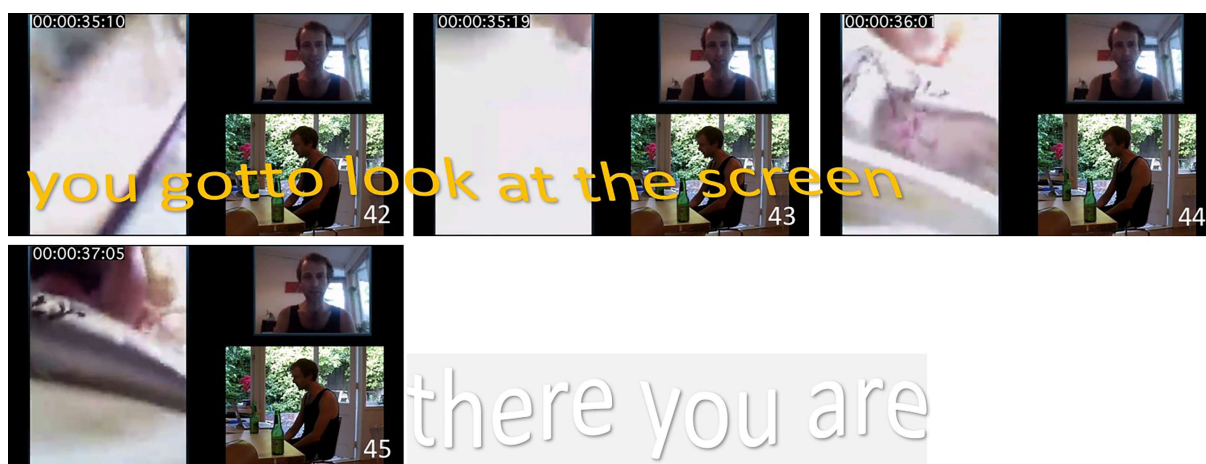


Figure 1: Mother instructing child where to look at the beginning of a Skype call.²

In images 42-45, we see the participant is sitting in front of the laptop, looking at the screen, listening to his sister speak with the child and then reacting to the child's face appearing on the screen. Here, it seems that we can easily determine in this brief excerpt that the participant is focused upon the Skype call with

¹ I would like to thank Freiburg Institute for Advanced Studies (FRIAS), University of Freiburg, Germany and the People Programme (Marie Curie Actions) of the European Union's Seventh Framework Programme (FP7/2007–2013) under REA grant agreement no. [609305] for making the writing of this article possible. I would also like to thank the Faculty of Design and Creative Technologies, the School of Communication Studies, and the AUT Multimodal Research Centre at Auckland University of Technology in New Zealand for funding the project that this article is based upon. Further, I would like to thank the participants in the Family Video Conferencing Interactions Project.

² Utterances: participant = white, partner = green, sister = yellow, researchers = pink, child = blue. All images are published with permission of the participants.

his sister and niece. He ostensibly demonstrates through his posture (which is positioned towards the laptop), his gaze (which is focused upon the screen), and his language use (listening and speaking), that he is focused upon the interaction. Particularly his use of language appears to be clearly indicating that he must be focused upon the conversation.

However, this article demonstrates that:

1. Focused attention can only be analyzed correctly when crossing micro-analytical boundaries. This means that if researchers cut their micro data pieces that they are investigating too short (as in the example in Figure 1), they are unable to detect and correctly analyze what participants are actually focused upon.
2. A participant can utilize language without paying focused attention to an interaction. Thus, if researchers assume that language use means focused attention, the assumption may in fact be incorrect.

The article builds upon a discussion of attention literature (Norris, forthcoming), which is not repeated here for space reasons, *and* revisits part of an excerpt that has been written about in Norris (2016), where an interaction around minute 4 of the video discussed here is analyzed. This article discusses the first two minutes in detail and illustrates how the focus of the participant is detected when taking a look at a data piece from the beginning of a recording. In Norris (2016, 152), it was demonstrated that it is not language which gives away what participants are focused upon, but rather modal density (of which language is a part) that comes about through modal intensity and/or modal complexity (Norris 2004). Here, the earlier point of that example illustrates the very beginning of the Skype call with a multimodal transcript (Figure 2-6) and honing in on the language that the participant uses (Audio Transcripts 1-4). Here, I show that the participant's use of language synchronization (not in a time-synchronized manner, but in a repetitive manner) allows the participant to fully function verbally even though he is *not focused* upon the interaction. Because I wish to demonstrate that a participant can smoothly interact verbally without being focused upon the interaction, I have chosen to represent these sequences in the form of both multimodal (Figure 2-6) and audio transcripts (Audio Transcript 1-4). A list of higher-level actions and the video excerpt discussed in this article can be found in Norris (2019, 190ff.; Video 5.2).

2. Data and data analysis

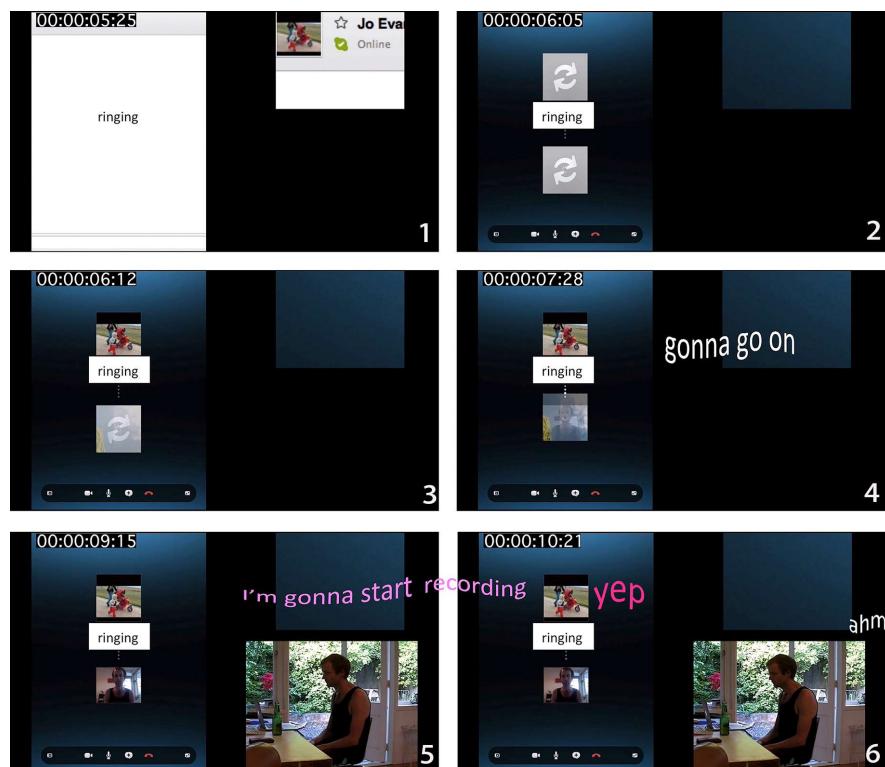
The data discussed in this article is part of a larger study of 17 New Zealand families and 82 participants from infants to an 80+ year old woman (mostly) Skyping with family members in Australia, Britain or Canada. Data collection occurred in the New Zealand participants' homes by one to three researchers at a time (depending upon availability) with a research laptop that had a screen recording software installed and one to two tripod-standing video cameras (depending on need and possibility) recording the interactions of the family members in the homes around the Skype interactions. A few weeks after the recording, (usually) a follow-up phone interview with at least one of the New Zealand adult family members was conducted (Norris 2019). The data in this article, however, comes from the video recorded data. This particular data piece was then multimodally transcribed following Norris's transcription conventions in order to ensure replicability and reliability and further analyzed in detail (Norris 2004, 2011, 2019).

Through a systematic and detailed analysis (Norris 2019), it becomes evident on the one hand that specific micro data pieces selected by researchers from a large amount of data, when a larger point of view is disregarded, can lead to incorrect or partial findings. When, on the other hand, micro-analytical boundaries are crossed, groundbreaking findings can be discovered and exact shifts in a participant's focused attention can be determined (Pirini 2014, 2015, 2017). In Norris (2016, 154f), it is shown that the participant shifts his focus to the Skype call between minute 3:54 and 3:57. Thus, rather than as argued in Norris (2011) that we need to employ a multimodal lens in order to gain greater insight into everyday interaction, this article demonstrates that it is *also the scale* of a data piece chosen (Norris 2017) that reveals lesser or greater insight into everyday interaction.

3. Beginning a Skype call

The data piece selected here shows a New Zealand participant during the initial Skype call to his sister in Australia. The data is recorded in his home on a research laptop and a tripod-standing camera. Two researchers are present and are interacting with the participant, the participant's partner and each other. Both researchers and the partner of the participant are out of camera view at this point. The partner is not audible at the very beginning, but then becomes audible and also partially visible in the video as she picks up a phone next to the participant in order to leave the room to call her mother (Figure 2). Then, later in the Skype conversation (not shown here), she also interacts with all Skype participants and becomes a participant herself.

Figure 2 is a multimodal transcript of the very beginning of the research session. The multimodal transcript follows multimodal transcription conventions (Norris 2002, 2004, 2011, 2019) and the utterances are color-coded to illustrate speaker changes. In Figure 2, image 1-5, we see the laptop screen as it is changing during the beginning of a Skype call. In image 4, we see the participant's utterance ›gonna go on‹ with the intonation pattern displayed as an approximate curve. Image 5 then shows that the first researcher says ›I'm gonna start recording‹ when the external tripod-standing camera begins to record the participant as he is sitting at a desk in front of the research laptop, trying to establish a connection with his sister. The second researcher responds to the first with ›yep‹ (image 6). A very brief moment later, the participant says ›ahm‹ (image 6) and continues (images 7-9) with ›am I sitting up straight‹. His voice starts out low and increases slightly in volume as he straightens up his posture (images 7-10) and as he turns to and looks at the researchers and his partner (images 9-10). In images 11-14, we see the participant turning back towards the laptop and with this turn, shifting his gaze and his head, slouching slightly forward, and relaxing his arms. Throughout, the participant demonstrates a wide smile, showing the humor in the question.



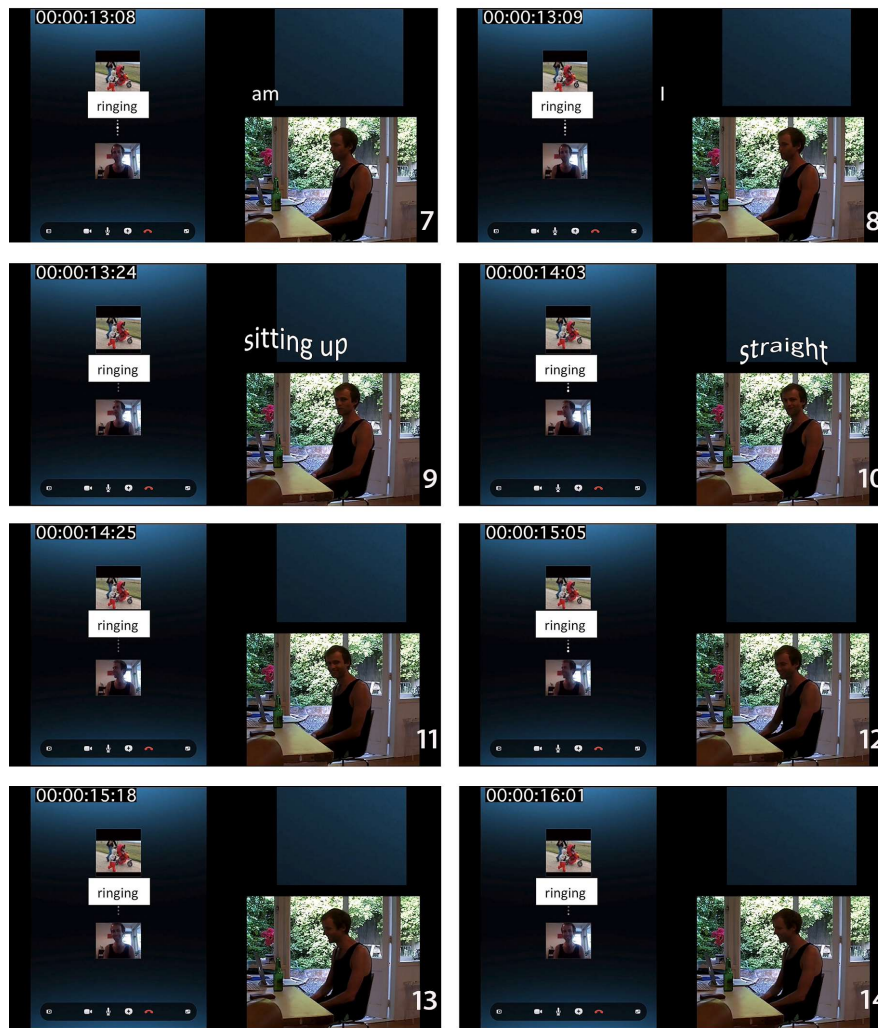


Figure 2: *Dialing up.*

What we see here is that the participant produces high modal density in his interaction with the researchers, showing his interactional focus. The modal density foreground-background continuum and, as mentioned before, the analysis of a slightly later excerpt are discussed in detail in Norris (2016). Here, the participant reacts to the researchers' utterances about starting the recording and makes a joke about sitting up straight for the camera and indicating this being funny through his demonstrative sitting up straight and smiling widely towards the researchers and the camera. Thus, the participant uses the mode of language, the mode of posture, the mode of hand-arm movement, the mode of head movement, the mode of gaze and the mode of facial expression, building up high modal density through both intensity (of language and facial expression) and complexity (through modal interconnectedness).

However simultaneously, as explicated in detail in Norris (2016, 152ff.), the participant is not unaware of the Skype call that he has initiated. Rather, he is paying medium attention to the call by sitting in front of the laptop, having his torso turned toward the screen so that he can easily be seen once the connection is established. He hears the ringing of the Skype call, and doubtlessly is listening to it. Further, we can surmise that he has not forgotten that his partner is in the room. They were engaged in interaction before the research session began, are engaged when she gets her phone and during technology breakdowns and are later interacting with his relatives together. Thus, we can say that even when the partner is not in the same room, the participant is aware of his partner if merely through proxemics (the partner being at home), paying some interactive attention to her.

4. Problematizing data piece selection

As discussed above, the participant is clearly focused upon the interaction with the researchers while he initiates his Skype call. However, we can only determine this focus when we include this segment in our analysis and transcribe this segment as illustrated in Figure 2. Yet, when a researcher dismisses this very segment as irrelevant and begins the analysis at a point when the participant is actually interacting with his sister and his niece(s) via Skype, the participant's actual interactional focus becomes obscured. In other words, when a researcher focuses only upon the actual Skype interaction as exemplified in Figure 1, the researcher is inclined to view the Skype conversation without hesitation as the participant's focused interaction. A participant's focus, I would like to argue, is most often not analyzed, rather it is usually presupposed by researchers in two respects:

1. The researcher's focus may be a Skype conversation or particular instances in Skype conversations such as an adult directing a child to look at the screen. Thus, a researcher may be interested in such interactions where participants on both sides of the screen are interacting with each other. Thus, the researcher presupposes that a participant focuses upon the interaction that the researcher is interested in.
2. The researcher presupposes that if a participant is engaged verbally with other participants, then the participant has to unquestionably be focused upon the interaction.

Here, I would like to argue that both of these presuppositions can be false and may lead to a misreading of focused interactions. As discussed in detail by Pashler (1998, 38), eye movement does not necessarily indicate a social actor's focused attention. Similarly, as discussed in detail by Norris (2011), language production does not necessarily indicate a social actor's focused attention.

According to the analysis in Norris (2016, 155ff.), the participant's shift in interactional focus occurs close to minute 4 in the data. Here is what happens: The piece transcribed above ends at 00:00:16:01. At 00:00:20:25, the Skype conversation begins with the adults greeting and an interaction emerges between the participant, his sister and one of her two daughters as discussed in detail below (see also Norris 2019, 190ff.). For about 105 seconds, the interaction runs smoothly, then a technology cut-off occurs. The participant and his partner interact during this cut-off. At minute 2, the connection is re-established, and the Skype conversation continues until a new technology glitch occurs around minute 3. Throughout these three minutes, the participant is focused upon either the researchers or his partner. Yet, he speaks with his sister in Australia and with two of her children (only interactions with one of them are discussed here). First, a multimodal transcript is presented and this is followed by an audio transcript. Audio transcripts use some conventions from Tannen (1984) so that: »?« means strong rising intonation, a comma means slight rising intonation, and a period means lowered intonation. Overlap is indicated with square brackets. The participant shown in Figure 1 is called Part (for participant), his partner in New Zealand is called Partner, Researcher 1 and 2 are R1 and R2 respectively, and the two children are called Child 1 and Child 2 in the audio transcripts. Further, the children's mother is here called Sister since she is the sister of the participant and our focus here is the participant.

Figure 3 is a direct continuation of the multimodal transcript in Figure 2 and Figure 1 is taken from the very last segment in Figure 3. The transcript (Figure 3) is then followed by the audio transcript (Audio Transcript 1), which demonstrates the language used by all in Figure 2 & 3. The language in the multimodal transcripts is color-coded (see footnote 2).

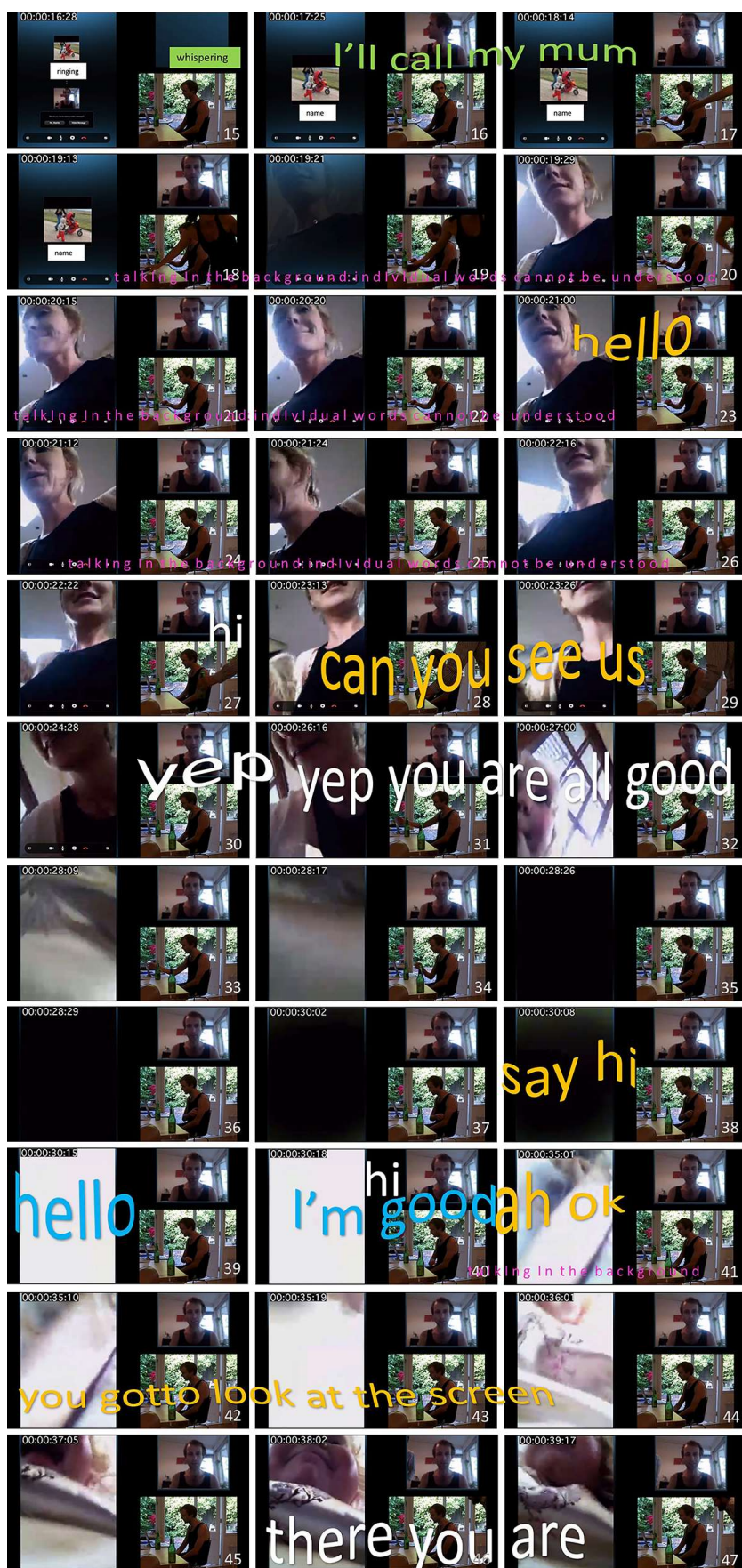


Figure 3: Connecting and beginning the interaction.

The first part (Audio Transcript I) begins at the same point as the multimodal transcript shown in Figure 2 and ends with the end of the multimodal transcript in Figure 3.

- (1) Part: gonna go on?
- (2) R1: I'm gonna start recording
- (3) R2: yep
- (4) Part: ahm
- (5) am I sitting up straight?
- (6) Partner: (whispering)
- (7) I call my mum
- (8) R2: (talking in the background: individual words cannot be understood)
- (9) Sister: hello
- (10) Part: hi
- (11) Sister: can you see us?
- (12) Part: yep yep
- (13) you're all good
- (14) Sister: say hi
- (15) Child 1: hello
- (16) Part: hi
- (17) Child 1: I'm good,
- (18) Sister: ah ok
- (19) you gotta look at the screen
- (20) Part: there you are,

Audio Transcript I: Beginning a Skype call.

As illustrated in Audio Transcript I lines 1 through 20, the participant is calling his sister in Australia, his partner in New Zealand is telling him that she is going to call her own mother and the two researchers are speaking quietly in the background (lines 2 & 3 and 8-11). As soon as the call goes through and the participant's sister picks up, the interactants greet each other (lines 9 & 10) and the sister immediately inquires about being seen. As soon as visual connection is assured, the sister asks one of her daughters to say hi (line 14) and the child does as she has been asked. Now the child greets the participant and the participant greets the child (lines 15-17). Then, the sister of the participant directs the child's gaze to the screen and the participant reacts to seeing the little girl's face on screen (lines 19 & 20; also Figure 1). Between the time the Skype call begins and the end of this excerpt, of which Figure 1 is a part, the participant fills 5 lines (Audio Transcript lines 9-20). However, what he says only requires medium attention on his part. The reason is that he uses Skype often and, according to our findings in the larger study, it is a most common opening to first greet each other and then inquire about whether one can be seen. Thus, here the participant speaks, going through the everyday motions when beginning a Skype call. His focus is still on being recorded for a research project even though he is engaged verbally with his sister and her young daughter. In Figure 3, image 26-29, we see a researcher placing his bottle of beer on the desk and in image 46 & 47 (which directly follows and actually overlaps with Figure 1), we see a researcher looking over the participant's shoulder. The proxemics of the researchers and his partner with the participant allow us to analyze the strong modal density that is produced between participant, researchers and partner. Thus, even though the participant appears to focus upon the Skype interaction through his posture, gaze and language use, he in fact is focused upon the interaction with the researchers (and at times with his partner). Here, in image 47 (Figure 3), we see the researcher smiling and right after (not shown here), the participant responds to the researcher's outbreath.

Next, the sister tells her brother that her daughter is dressed up for him and the conversation continues about what she is wearing. Here, the participant utilizes common social etiquette asking what the girl is wearing and then commenting that it is lovely. Again, the participant's focused attention is not needed to continue the conversation and appear fully engaged. Here, I say »appear engaged«, because later, almost at the end of the three minutes, we see that his sister in fact knows that he is not fully focused upon the

conversation as she inquires ›are you alright?‹ But there are other ways than social etiquette and regular openings of a Skype conversation that he utilizes to engage in a conversation without being fully focused upon it. He does this by synchronizing his utterances to the interlocutors' through repetition.

For some time, the conversation is driven by the little girl, who speaks with the participant about coming to his house (Figure 4 and Audio Transcript 2).

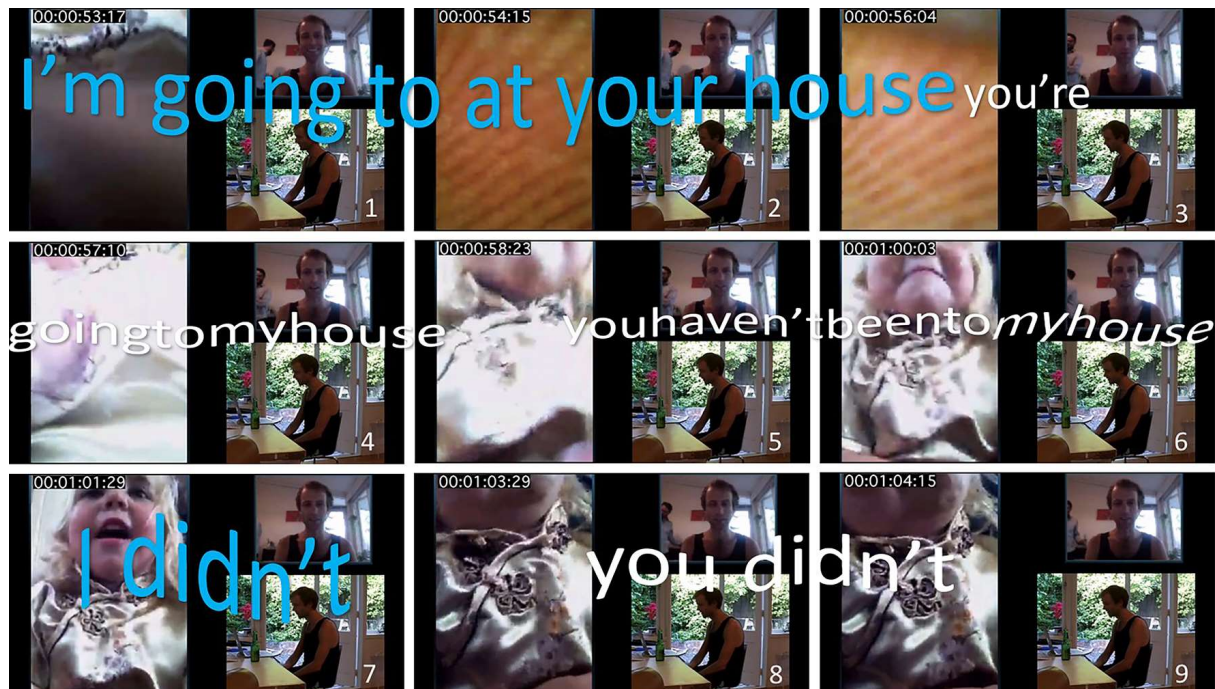


Figure 4: Three-year old conversing with the participant (her uncle).

At her young age of 3, she tries to express herself as shown in Figure 4 and in Audio Transcript 2 (lines 33-37).

- (33) I'm going to at your house
 (34) Part: you going to my house,
 (35) you haven't been to my house?
 (36) Child 1: I didn't!
 (37) Part: you didn't

Audio Transcript 2: Synchronizing utterances with those of the child.

But what is remarkable here is that the participant uses the child's utterances, reformulates them slightly and parrots back to the child what she has said with only very slight changes (Figure 4 and Audio Transcript 2).

Then, when for example his sister comes to the rescue to clarify what the little girl was telling the participant (Figure 5, image 7 and Audio Transcript 3, line 54), he again uses repetition to continue the conversation without much attentional effort. Here, the child tells her uncle what she wants to get, her mother corrects the word ›Mohawk‹ and the participant repeats part of the word, ending in OK.



Figure 5: Mother correcting speech.

In all nine images of the multimodal transcript (Figure 5), we see a researcher standing or moving behind the participant. Clearly, the participant, who can see the researcher on the laptop screen, is highly aware of being recorded and observed. However, he engages continuously with his family members via Skype. He does this easily through the use of repetition (Audio Transcript 3).

- (48) Child 1: ah
 (49) I
 (50) I'm gonna get a
 (51) I'm gonna get a
 (52) spider cake and a mohkak.
 (53) Part: spider cake
 (54) Sister: and a mohawk
 (55) Part: and a mohawok

Audio Transcript 3: Synchronizing utterances with the child and the sister.

Thus, the participant skillfully repeats what the child and his sister say by using their utterances and forming them into his own (Bakhtin 1981). By doing so, he appears to be listening and paying close attention without a need to actually pay focused attention. Social etiquette and synchronization of his utterances with those of the interlocutors thus enable him to verbally engage in the mid-ground of his attention, while he is simultaneously still focused upon the research session. Synchronization in interaction has been shown to produce connection (Breyer et al. 2017). Verbally synchronizing, not at the time when the same thing is being said, but *synchronizing what is being said*, allows the participant in the above example to establish and display connection while he mid-grounds this interaction. As shown earlier (Norris 2011), in close relationships where one interlocutor pays focused attention to the other (here, it would be the little girl, for example), while the other attends to the conversation in the mid-ground (here, it would be our participant), enables the focused interlocutor to speak freely. Thus, depending upon the interaction, such an attention constellation, where one interlocutor pays focused attention and the other mid-grounds the interaction, can be experienced as comfortable by the person paying focused attention.

The fact that the participant is not fully focused upon the Skype interaction can also be seen when his partner chimes in and comments on the little girl as soon as a technology breakdown occurs. At that

point, this brief exchange occurs (Multimodal Transcript (Figure 6) and Audio Transcript 4, lines 61-67).

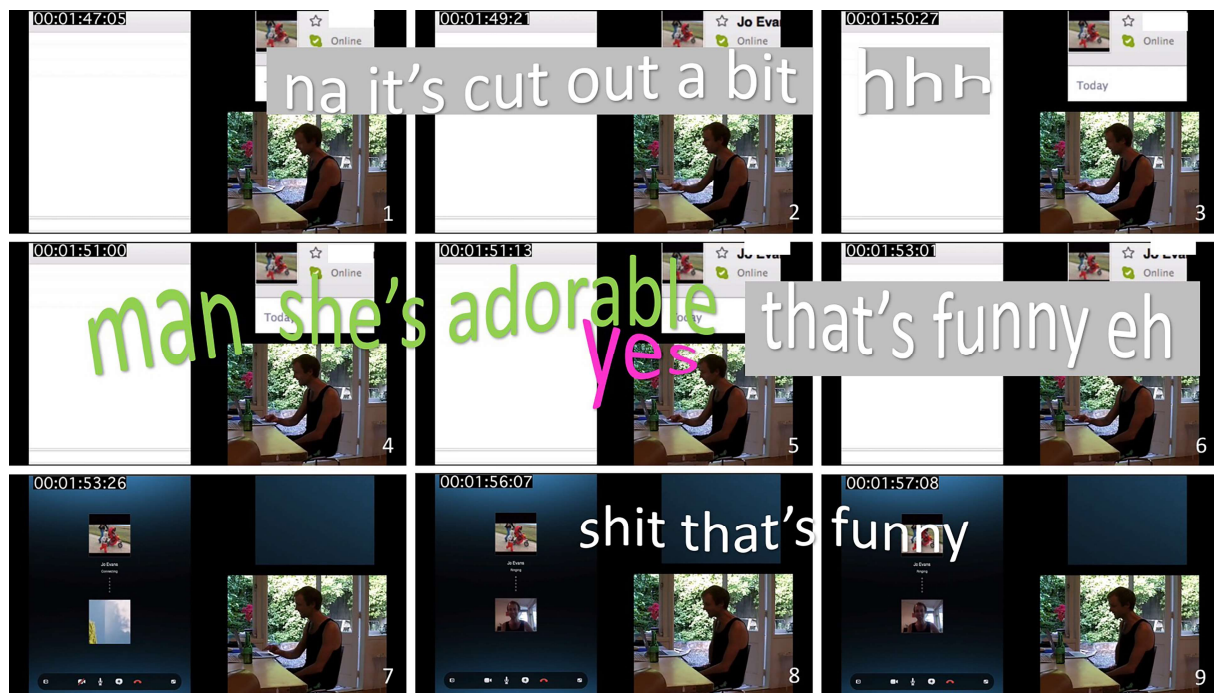


Figure 6: Connection is lost.

- (61) Part: na it's cut out a bit
 (62) hhh
 (63) Partner: man!
 (64) she's adorable,
 (65) R 2: yes
 (66) Part: that's funny eh,
 (67) shit that's funny.

Audio Transcript 4: Conversation between participant and his partner.

However, in Figure 6 (images 5 & 6 and 8 & 9) and Audio Transcript (lines 65-67), the participant sounds thoughtful rather than with a gist of humor as one would expect when a person speaks about something being funny. As for example, demonstrated in Helmholtz (1896) or Cherry (1953), attention is selective and here, it appears that the participant is actively selecting to shift his focus to the Skype interaction. This shift, however, does not occur immediately. As shown in Bernad-Mechó (2017), a shift in focus often goes through an intermediate stage in which the interlocutor is neither fully focused upon one interaction nor on the other. When the participant reconnects, he becomes more active as an interlocutor, asking about the other girl and speaking about her hair. But in the next few utterances, he again reverts to synchronizing his utterances with those of his interlocutors' previous utterances and then again becomes more active as he asks questions of the girls. The point here is that he seems to be going through a transition from focusing upon taking part in a research project and mid-grounding the Skype interaction to focusing upon the Skype interaction and mid-grounding the research session. His sister reacts to his *interactional attention* when she asks ›are you alright?‹ and briefly after, as discussed in detail in Norris (2016, 160), he refocuses completely and focuses upon the Skype session. As illustrated there, the participant displays a clear multimodal focus so that he uses his posture, his facial expression, hand/arm movements and gaze as well as language to interact via Skype. His engagement has changed not only multimodally, but also in the rhythm of his speech. Thus, we find a change in rhythm once he has refocused. While the rhythm in the first three minutes of the Skype call is slow, the rhythm increases in intensity as soon as the participant has refocused.

5. Conclusion

This article problematizes analytical approaches to discourse, where history, memory, and overall context are disregarded, where minute samples are cut from much larger data pieces without taking a broader view of the data and where researchers or cameras recording the participants are assumed to be irrelevant to the participants' attention. Participants' attention can neither be presupposed based on micro-data piece selection by a researcher, nor can it be presupposed based on language use by the interlocutors. In order to make this point, the article began by first showing a micro-data piece (Figure 1), where the focused interaction of the participant seems to be apparent. Then, by analysing the few images from Figure 1 in their context (Figure 3, images 42-45), it is demonstrated that micro-data pieces easily are misleading. Figure 1, as shown above, was a micro excerpt taken from the end of the multimodal transcript (Figure 3), which clearly demonstrates that, when looking at the longer excerpt, the participant is actually *not focused* upon the Skype call in Figure 1.

Current and past attention literature gives insight into aspects of attention and is explicated in Norris (forthcoming). But briefly, scholars differ in their assumption of what happens during interaction. Gundel et al. (1993), Brennan (1995), Levelt (1989), or Clark and Marshall (1981), for example, work with the theoretical assumption of ›speakers' models of listeners' knowledge‹ (Bard et al. 2000, 3). At a later moment in the interactions (which I can only touch upon here), we find that the sister (in that case the speaker) reacts to the displayed mid-grounded attention by the participant (in that case the hearer) when she asks whether he is all right. Thus, here, we find a case of the speakers' model of the listeners' attention. Simultaneously, however, we find the opposite viewpoint posited by scholars such as Chafe (1994), Arnold and Lao (2015), or Bard et al. (2000), who claim that speakers focus more on their own speech than on the interlocutor. This may be particularly evident when the participant selects to pay focused attention to the Skype call rather than to the research session. Even though this selection takes time to fully perform, he changes his speaking style to ask questions and become more actively involved.

This article thus demonstrates *interactional attention* (Norris 2002, 2004, 2006, 2008, 2011, 2016, 2019), which is an aspect of a phenomenal conception of attention. Interactional attention (Norris 2004) converges the two points of view on attention discussed above. In other words, with this framework, we see that interlocutors do judge and react to the interactional attention of others. At the very same time, with this framework, we can determine when and how speakers focus more on their own than on others' actions. In order to properly analyze interactional attention, we have to analyze each interlocutor individually *and* analyze the interlocutors together, as suggested in Norris (2011). The reason for this is that one interlocutor may pay a different level of attention to an interaction than another (see also Norris 2006).

This article furthermore demonstrated that synchronization of utterances, not at the same time, but in form of repetition, can allow an interlocutor to interact verbally without having to pay focused attention. This kind of synchronization can have two effects (depending upon the situation): 1. It can demonstrate that the one synchronizing their utterances to those of the other is listening; and 2. It can produce a connection. However, while this is the case in the interaction analyzed above, more research is needed to determine under which circumstances such synchronization functions in this way. Further, it has been suggested above that the rhythm of interaction changes when a person focuses upon it after having previously mid-grounded the interaction. Here too, more research is needed to discover if a change of rhythm is always present after a change in focus.

The article thus shows that when theorizing attention as *interactive attention*, we can determine the attention level of participants in interaction in a theoretically grounded manner. This way of working entails a broader point of view of studying interaction, both in the direction of modal use besides language (i.e. multimodality) and in the direction of delineating what it is that we are examining (moving the research interest beyond the instance that researchers may find relevant). Modal density, it is shown, is achieved through either intense or complex usages of modes, resulting in rhythm and pace of speech as well as rhythm and pace of other modes (Norris 2009).

This article critically assesses the assumption that language use by participants necessarily leads the researcher to the participants' focused interaction. As shown with the multimodal transcripts in Figures I-5 and in connection with this in Norris (2016), a true focus of attention by a participant can only be determined if we take a broader view of our data. If we are simply picking and choosing brief excerpts that we as researchers are for whatever reason focused upon, we cannot make any claims about the focus of the participants. Further, we can make no claim about the relevance or importance of the language that is being used. Language may be used by a participant in the focus, but language may also be used in the mid-ground and even in the background of a participant's attention. The study of interactional attention promises to help us discover a whole host of new findings that, as long as we as researchers insist that language production always occurs in the focus of a participant, we may actually miss. Thus, we need to stop picking minute interactional sequences which disregard the bigger data pieces that the minute ones are a part of.

Here, it is necessary to realize that whether language is utilized in the focus, the mid-ground or the background of somebody's always multimodally displayed attention does not make language less important! Rather, it is specifically the fact that social actors can and do use language on a range of attentional levels that moves us into a new and highly promising direction of research. Norris (2011), for example, showed that a participant, who was focusing upon a conversation with her friend, was delighted that her friend did not reciprocate the attention that she herself paid to the conversation. In fact, the participant who was focusing upon the conversation felt safe, un-judged and taken seriously by her friend who was mid-grounding the conversation. Similarly, school children may wrongfully be told to pay focused attention and look at their teacher, when in fact they might be learning much better by not displaying interactional focus to what is being said or done. Similarly, other interactions where the one in a lower power position is asked to give information such as some manager-worker interactions or some parent-child interactions and possibly even some doctor-patient interactions may proceed much smoother if the one in power does not pay focused interactional attention to the interlocutor to demonstrate the one giving the information is safe, un-judged and taken seriously. However, this is only a suggestion and much research is needed in order to determine how and when interactional focus is helpful and when it is not. But one thing is certain: Research into *interactional attention*, which is research that crosses micro-analytical boundaries, will have social ramifications with practical dimensions.

References

- Bakhtin, Mikhail Mikhailovich 1981: *The Dialogic Imagination*. Austin, TX: University of Texas Press.
- Bernad-Mechó, Edgar 2017: Metadiscourse and Topic Introductions in an Academic Lecture: A Multimodal Insight. In: *Multimodal Communication* 6/1, 39-60.
- Breyer, Thimo/Buchholz, Michael/Hamburger, Andreas/Pfänder, Stefan (eds.) 2017: *Resonanz, Rhythmus & Synchronisierung: Interaktionen in Alltag, Therapie und Kunst*. Bielefeld: transcript.
- Chafe, Wallace. L. 1994: *Discourse, consciousness, and time*. Chicago: Chicago University Press.
- Clark, Herbert H./Marshall, Catherine R. 1981: Definite reference and mutual knowledge. In: Aravind K. Joshi/Bonnie L. Webber/Ivan A. Sag (eds.): *Elements of discourse understanding*. Cambridge: Cambridge University Press, 10-63.
- Cherry, E. Colin 1953: Some experiments on the recognition of speech, with one and with two ears. *Journal of the Acoustic Society of America*. (25): 975-979.
- Helmholz, Hermann von 1896: *Handbuch der physiologischen Optik*. L. Voss.
- Levelt, Willem J. M. 1989: *Speaking*. Cambridge: MIT Press.
- Norris, Sigrid 2002: *A theoretical framework for multimodal discourse analysis presented via the analysis of identity construction of two women living in Germany*. Dissertation. Department of Linguistics, Georgetown University.
- Norris, Sigrid 2004: *Analyzing Multimodal Interaction: A Methodological Framework*. London: Routledge.
- Norris, Sigrid 2006: Multiparty interaction: a multimodal perspective on relevance. *Discourse Studies* 8/3, 401-421.

- Norris, Sigrid 2009: Tempo, Auftakt, levels of actions, and practice: rhythms in ordinary interactions. *Journal of Applied Linguistics* 6/3, 333-356.
- Norris, Sigrid 2008: Some thoughts on personal identity construction: A multimodal perspective. In Bhatia, Vijay, Flowerdew John, and Jones, Rodney, H. (eds) *New Directions in Discourse*. London: Routledge. 132-149.
- Norris, Sigrid 2011: *Identity in (Inter)action: Introducing Multimodal (Inter)action Analysis*. Berlin/Boston: Mouton.
- Norris, Sigrid 2017: Scales of action: An example of driving & car talk in Germany and North America. *Text & Talk*. 37(1): 117-139.
- Norris, Sigrid 2019: *Systematically working with multimodal data: Research methods in multimodal discourse analysis*. Hoboken, NJ: John Wiley and Sons.
- Norris, Sigrid (Forthcoming): *Multimodal Theory and Methodology: for the Analysis of (Inter)action and Identity*. New York: Routledge.
- Pashler, Harold. E. 1998: *The psychology of attention*. Cambridge, MA: MIT Press.
- Pirini, Jesse 2014: Producing Shared Attention/Awareness in High School Tutoring. In: *Multimodal Communication* 3/2, 163-179.
- Pirini, Jesse 2015: *Tutoring as Knowledge Communication: A multimodal (Inter)action Analysis*. Unpublished PhD thesis. Auckland University of Technology, New Zealand.
- Pirini, Jesse. 2017. Agency and Co-production: A Multimodal Perspective. *Multimodal Communication* 6/2, 1-20.
- Tannen, Deborah 1984: *Conversational Style: Analyzing Talk Among Friends*. Norwood, NJ: Ablex.

Online sources

- Arnold, Jennifer E./Lao, Shin-Yi C. 2015: Effects of psychological attention on pronoun comprehension. In: *Language, Cognition and Neuroscience* 30, 832-852. DOI: [10.1080/23273798.2015.1017511](https://doi.org/10.1080/23273798.2015.1017511).
- Bard, Ellen Gurman/Anderson, Anne H./Sotillo, Catherine/Aylett, Matthe/Doherty-Sneddon, Gwyneth/Newlands, Alison 2000: Controlling the intelligibility of referring expressions in dialogue. In: *Journal of Memory and Language* 42 (1), 1-22. DOI: [10.1006/jmla.1999.2667](https://doi.org/10.1006/jmla.1999.2667).
- Brennan, Susan E. 1995: Centering attention in discourse. In: *Language and Cognitive Processes*, 10(2), 137-167. DOI: [10.1080/01690969508407091](https://doi.org/10.1080/01690969508407091).
- Gundel, Jeanette K./Hedberg, Nancy/Zacharski, Ron 1993: Cognitive status and the form of referring expressions in discourse. *Language* 69, 274-307. DOI: [10.2307/416535](https://doi.org/10.2307/416535).
- Norris, Sigrid 2016: Concepts in multimodal discourse analysis with examples from video conferencing. *Yearbook of the Poznań Linguistic Meeting* 2 (1), De Gruyter Open, 141-165. ISSN (Online): 2449-7525, DOI: [10.1515/ypplm-2016-0007](https://doi.org/10.1515/ypplm-2016-0007).