Analyzing Multimodal Interaction

“This is an outstanding book, strikingly well focused to the point of view of readers new to its concepts. It will be essential, required reading on many courses and will be very useful in opening up a major cross-disciplinary literature and approach.”

Ron Scollon, Georgetown University, USA

“The wide, varied range of materials discussed, matched with precise, detailed methods of analysis, demonstrated by carefully worked examples, will make this indispensable for anyone interested in getting at meaning in its rich variety.”

Gunther Kress, Institute of Education, University of London, UK

Our perception of our everyday interactions is shaped by more than what is said. From coffee with friends, to interviews; and from meetings with colleagues, to conversations with strangers, we draw on both verbal and nonverbal behavior to judge and consider our experiences.

Analyzing Multimodal Interaction is a practical guide to understanding and investigating the multiple modes of communication, and provides an essential guide for those undertaking fieldwork in a range of disciplines, including linguistics, sociology, education, anthropology, and psychology. The book offers a clear methodology to help the reader carry out their own integrative analysis, equipping them with the tools that they need to analyze a situation from different points of view. Drawing on research into conversational analysis and nonverbal behavior such as body movement and gaze, it also considers the role of the material world in our interactions, exploring how we use space and objects – such as our furniture, cell phones, or TV. Considering a range of real examples, such as traffic-police officers at work, doctor–patient meetings, teachers with students, and friends reading magazines together, the book offers lively demonstrations of multimodal discourse at work.

Illustrated throughout and featuring a mini-glossary in each chapter, further reading, and advice on practical issues such as making transcriptions and video and audio recordings, this practical guide is an essential resource for anyone interested in the multiple modes of human interaction.

Sigrid Norris teaches in the Department of English, Linguistics, and Speech at Mary Washington College, Virginia, USA. Her research areas include multimodal identity construction and multimodal cross-cultural communication.
Analyzing Multimodal Interaction
A methodological framework

Sigrid Norris
To
Alan, Kevin, and Luke
Contents

List of figures viii
List of plates ix
Preface x
Acknowledgments xii

1 Multimodal interaction 1
2 Communicative modes 11
3 Multimodal transcription 58
4 Modal density 79
5 Levels of attention/awareness 95
6 Semantic/pragmatic means 112
7 Modal density foreground–background continuum as a methodological framework: complex interactions 128
8 Analyzing multimodal interaction: a postscript 148

References 156
Index 174
Figures

5.1 Modal density foreground–background continuum 99
5.2 Modal density circles as heuristic tools to visualize the
density employed for specific higher-level actions 108
5.3 Sandra’s six simultaneously constructed higher-level
actions placed on a modal density foreground–background
continuum 110
## Plates

<table>
<thead>
<tr>
<th>Plate</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.1</td>
<td>A crossing guard communicating on different levels of attention/awareness</td>
<td>8</td>
</tr>
<tr>
<td>2.1</td>
<td>Proxemics</td>
<td>22</td>
</tr>
<tr>
<td>2.2</td>
<td>Posture</td>
<td>27</td>
</tr>
<tr>
<td>2.3</td>
<td>Gesture</td>
<td>31</td>
</tr>
<tr>
<td>2.4</td>
<td>Head movement</td>
<td>35</td>
</tr>
<tr>
<td>2.5</td>
<td>Gaze</td>
<td>39</td>
</tr>
<tr>
<td>2.6</td>
<td>Music</td>
<td>43</td>
</tr>
<tr>
<td>2.7</td>
<td>Print</td>
<td>47</td>
</tr>
<tr>
<td>2.8</td>
<td>Layout</td>
<td>50</td>
</tr>
<tr>
<td>2.9</td>
<td>Interconnection of modes</td>
<td>55</td>
</tr>
<tr>
<td>3.1</td>
<td>Transcript: proxemics and posture</td>
<td>70</td>
</tr>
<tr>
<td>3.2</td>
<td>Transcript: gesture</td>
<td>71</td>
</tr>
<tr>
<td>3.3</td>
<td>Transcript: head movement and gaze</td>
<td>73</td>
</tr>
<tr>
<td>3.4</td>
<td>Multimodal transcript</td>
<td>76</td>
</tr>
<tr>
<td>4.1</td>
<td>High modal density: layout and music</td>
<td>85</td>
</tr>
<tr>
<td>4.2</td>
<td>Modal density: high density</td>
<td>86</td>
</tr>
<tr>
<td>4.3</td>
<td>Modal density: modal complexity</td>
<td>89</td>
</tr>
<tr>
<td>4.4</td>
<td>Modal density: intensity plus complexity</td>
<td>90</td>
</tr>
<tr>
<td>5.1</td>
<td>Transcript: six simultaneous higher-level actions</td>
<td>102</td>
</tr>
<tr>
<td>6.1</td>
<td>Semantic/pragmatic <em>means</em>: deictic leaning towards an object/participant</td>
<td>119</td>
</tr>
<tr>
<td>6.2</td>
<td>Semantic/pragmatic <em>means</em>: eyebrow flash</td>
<td>123</td>
</tr>
<tr>
<td>6.3</td>
<td>Semantic/pragmatic <em>means</em>: hand/finger beat</td>
<td>125</td>
</tr>
<tr>
<td>7.1</td>
<td>Transcript: music instruction</td>
<td>130</td>
</tr>
<tr>
<td>7.2</td>
<td>Construction of social worlds</td>
<td>139</td>
</tr>
<tr>
<td>7.3</td>
<td>Transcript: multimedia and interaction</td>
<td>144</td>
</tr>
</tbody>
</table>
Preface

This book is an attempt to explicate a methodological framework for the analysis of human interaction in its vast complexity. With a foundation in discourse analysis, interactional sociolinguistics, mediated discourse analysis, and multimodality, I cross the boundaries between linguistics, nonverbal behavior, and the material world.

Language is of particular interest to the study of human interaction, as the mode has great informative as well as expressive value. This mode is probably the best understood mode so far. Yet people in interaction seldom communicate only through language. A person takes up a certain kind of distance to others, takes up a particular posture, gestures while speaking, and at times gazes at the interlocutor.

Modes like gesture, gaze, or posture have generally been termed nonverbal modes of communication. However, I will steer away from this expression, as nonverbal conveys that these are appendages to the verbal mode. If the so-called nonverbal modes were actually appendages to language, these modes would always have to be subordinate to language. However, this is not the case. Modes like gesture, gaze, or posture can play a superordinate or an equal role to the mode of language in interaction, and therefore, these modes are not merely embellishments to language. In this book, I only refer to these modes as nonverbal when I want to point out which generally accepted fields of study I am working in. But, exclusive of this, I use the term embodied modes: a term that refers to gesture as well as to language, showing that the modes are generally of an equal value, and allowing the analyst to decide which mode (if any) plays a superior role in a particular interaction.

All interactions take place in the material world, and the material world plays a role in every interaction. With material world I do not only mean the setting that the interaction takes place in, but also the material world that people in interaction utilize. People are dressed in a certain way, they eat, they listen to music, and they read magazines. All of this may be a part of an interaction. I call modes like music, print, or layout disembodied modes. These modes can also take on a superordinate role in interaction, and they can “overrule” embodied modes.
Thus, we have to consider all embodied modes (like language, gesture, and gaze), and all disembodied modes (like music, print, and layout) that people in interaction utilize. Additionally, we cannot think of human interaction without taking the human mind into consideration. A person always thinks, perceives, and/or feels something when interacting with others, and at least some of these thoughts, perceptions, and/or feelings are communicated through a person’s actions. In turn, these actions trigger thoughts, perceptions, and/or feelings in the other person, who then also reacts by communicating at least some of their own thoughts, perceptions, and/or feelings through their own actions.

Although these considerations point to a theoretical basis, I actually devised the methodological framework for multimodal interactional analysis out of a practical need. A few years ago, I set out to ethnographically study the everyday identity construction of two women living in Germany. My findings were clear: the women constructed their everyday identity fragments on multiple levels in multiple modes. While I could talk about the findings, there was no applicable framework that allowed me to show scientifically how the women constructed several identity fragments simultaneously by employing different communicative modes. As a result, I developed the framework that I explicate in this book. Thus, this framework for multimodal interactional analysis is a practice-based methodology.
Acknowledgments

Many people have contributed to this book in various ways. I would like to particularly express my gratitude to Ron Scollon for the hours of discussion, his encouragement, and comments on the manuscript; Gunther Kress for his insightful comments; and Heidi Hamilton, Theo Van Leeuwen, and Ruth Wodak for their fruitful conversations. I spoke with numerous people at universities, conferences, and also via email about my framework, and all of these people had some impact on the development of my thoughts. Although I shall not attempt to name each person here, I would like to thank particularly Najma Al Zijdaly, Cecilia Castillo-Ayometzi, Ingrid de Saint-Georges, Doug Harper, Allison Jablonko, Andy Jocuns, Alex Johnston, Rodney Jones, Philip LeVine, David McNeill, Marilyn Merritt, Maurice Nevile, Jon Prosser, Suzie Scollon, and Jon Wagner. I am grateful to Tom Randolph for his detailed attention to wording.

A multimodal framework for the analysis of multimodal interaction can only be explicated multimodally. Images are as much a part of this book as the written words, and I would like to thank Maya Betts, Tracy Betts, Andrea Brandt, Annette Brandt, Gloria Buchanan, Shirley A. Hanson, Robert Le May, Marilyn McGonegle, Mary Lou Nale, Alan G. Norris, Kevin P. Norris, Luke Norris, Graziella Röntgen, John Schucker, Tammy L. Shapiro, and Suzanne S. Williams, DMD, who generously agreed to be videotaped and to have their pictures incorporated in the book to help me explicate the methodological framework.

My family has taken part in the book in countless ways, some visible and others not visible in the book at all. My sons, Kevin and Luke, have spent hours reading in the same room that I was writing in, and at other times they have made sure that I escaped from my desk. My partner, Alan, has spent hours reading and rereading the manuscript, giving me comments about wording. The times that Kevin and Luke spent sitting close by were most productive for me – those fun times were always refreshing, and Alan’s constant loving support was most valuable: it is to them that I dedicate this book.

While I am indebted to all of these people, as well as many others who helped in my thinking, the statements made and the views expressed are solely the responsibility of the author.
7 Modal density foreground–background continuum as a methodological framework

Complex interactions

Summary: The modal density foreground–background continuum explicitly links interaction to consciousness, showing that participants can construct several higher-level actions simultaneously – each of which is sequentially structured.

This framework allows us to qualitatively analyze the interplay of embodied and disembodied communicative modes, and we no longer have to differentiate between face-to-face and mediated interaction.

By qualitatively analyzing the multiple modes that participants employ, we recognize that any mode may take on high intensity and thus take on the hierarchically supreme position, structuring the other modes utilized. On the other hand, no mode has to necessarily be hierarchically superior to other modes within a specific interaction.

Complex multimodal interactions are structured by a "communicative semantics/pragmatics" or semantic/pragmatic means. Participants structure their own foregrounding of higher-level actions through the performance of means, and simultaneously indicate an upcoming shift in their foregrounded higher-level action to other participants.

Means can also be specifically used to influence another participant, indicating to a participant that they are expected to follow a shift in foregrounded higher-level action.

In this chapter, I focus on practical examples of day-to-day interactions, to illustrate how this methodology can lead to some new insights about human interaction. Here I draw on some new examples, but also return to some examples from previous chapters.
Multimodal teaching:音乐教学

### Multimodal teaching:
A teacher can utilize modes that are usually used interdependently (like gesture and spoken language) as two distinct meaning making systems, when such a distinction is beneficial for teaching/learning.

### Utilizing a means:
A teacher can “pull” the students’ focused attention towards the teacher’s newly foregrounded higher-level action by employing a means.

Instructing a student to learn a musical instrument is a complex and multimodal process. Both teacher and student(s) employ multiple modes to construct the higher-level action of a music lesson. As soon as a teacher and student(s) are fully engaged in the process, the higher-level action of giving/receiving a music lesson is backgrounded, and several other higher-level actions are foregrounded and mid-grounded.

As explained in Chapter 5, the actual hierarchical level of the higher-level actions performed is not too important to the participants, as they keep multiple actions — no matter which hierarchical level — on some level in their consciousness. Consequently, when a teacher instructs a student how to play the guitar, both participants are aware of this higher-level action throughout the interaction — in the background of their attention/awareness.

The following example illustrates the complexity of a music lesson. The teacher is utilizing multiple modes to facilitate learning, and Peter utilizes multiple modes to acquire a new strumming pattern.

The first image of Plate 7.1 shows the guitar teacher lifting his right hand while he says “up,” and the second image of the first row illustrates that he lowers his right hand while he says “down.” He performs such chains of lower-level actions 19 times between 7:28.46 and 8:03.35. However, while the chain of lower-level actions of the hand/arm movement does not change, the chain of the lower-level actions of saying “up” and “down” is not concurrent with all the movements of the teacher’s arm. The teacher utilizes the mode of spoken language, saying “up” or “down,” to indicate that Peter has to stroke the strings at those times. Whenever the teacher only moves his hand/arm up or down without uttering the co-occurring words, Peter is not supposed to touch the strings, but move his hand up and down to keep the rhythm and ensure that his hand is in the right position when he comes to the next note that requires strumming. During the 19 iterations of hand/arm movements of the teacher, he only utilizes co-occurring utterances, saying up or down 11 times.
Plate 7.1 Transcript: music instruction.
The hand/arm gestures of the teacher are used iconically to depict the movement of the student’s hand/arm or, in other words, the gestures convey through pictorial content the correct hand/arm movements that Peter is supposed to employ. Thus, the teacher utilizes these iconic hand/arm gestures to communicate and facilitate the movement, scaffolding to help Peter perform the regular movements that are necessary for the new strumming pattern. Concurrently, the teacher utilizes the mode of spoken language to communicate the voiced part of the strumming pattern to his student, again scaffolding to help the student to perform the new strumming pattern.

Here, the two modes are sometimes used together, and sometimes only the mode of hand/arm gesture is employed. However, during the times when gesture and spoken utterance are employed simultaneously, these two modes actually convey two different concepts. The hand/arm gesture always and only communicates the correct rhythmic movement of the hand/arm to the student, while the spoken utterances always and only communicate the actual strumming of the strings to Peter. Thus we see that the two modes, which sometimes co-occur, are used by the teacher and understood by the student as separate meaning-making resources.

During 17 up/down hand/arm movements, the teacher and his student are in synchrony. They are both utilizing the modes of gaze and print: reading the strumming pattern from a book that is placed on the music stand, as well as the mode of hand/arm movement and music. The teacher’s hand/arm movements are prominent, and Peter can follow these movements in his peripheral vision. Once Peter has performed 17 movements, he becomes confused, and he stops playing after performing two more movements, as his teacher keeps making hand movements. For the last two movement phases, Peter and his teacher are no longer in synchrony. While the teacher keeps the rhythm, Peter lags slightly behind in his movements.

As soon as the teacher stops his up/down movements, he performs a means; at 8:03.35 the teacher points to Peter with a quick deictic hand movement (as illustrated in the first image of the second row in Plate 7.1) and says “OK, also another thing.”

This means indicates a shift in higher-level action, from practicing the new strumming pattern with his student to verbal teaching. The guitar teacher utilizes the means pragmatically causing his student to refocus on the teacher’s newly foregrounded higher-level action. Once the teacher has performed the means, he points to the book on the music stand (as illustrated in the second image of the second row in Plate 7.1) and says “if you accidentally miss one of the strokes.” Here he starts turning his head towards Peter. The first image in the third row shows that the teacher is now looking at Peter while he continues to speak saying “that you’re supposed to hit.” Here, Peter is gazing at the teacher’s index finger, but, only a second later, Peter has shifted his gaze towards the teacher (as illustrated...
in the second image of this row). At this point, the teacher performs four quick back-and-forth hand movements that resemble erasing something, as illustrated by the arrow and the number four in the second image in the third row. While the teacher performs this iconic gesture, he says “don’t go back.” The teacher then points at the pattern in the book (as illustrated in the first image of the last row) saying “and you just play through.” While continuing his utterance, saying “and make sure you catch it the next time around,” he moves his hand back into a rest position, gazing at Peter (as illustrated in the last image of Plate 7.1).

**Multimodal methodology: independence or interdependence of modes**

This example demonstrates that even modes that are usually interdependent – here, gesture and language – can be utilized simultaneously, yet independently. The up/down movements of the teacher’s right hand/arm and the utterances “up” and “down” build two distinct chains of lower-level actions: being utilized as two distinct meaning-making resources by the teacher and the student. Between 7:28.46 and 8:03.35, the modes of gesture and spoken language are distinct and of equal value when trying to hierarchize the modes. Each mode is utilized to communicate a separate aspect of the strumming pattern, and both are of equal importance.

Only when we think of communicative modes as heuristically distinct chains of lower-level actions, can we realize when the modes are used as distinct modes and when they are used in a highly interdependent way.

As we saw in the second image of the third row in Plate 7.1, this iconic gesture that the teacher performs is highly dependent upon the utterance “don’t go back.” This gesture could not be understood properly by the student, without the teacher’s use of the mode of spoken language. Therefore, this gesture is dependent upon the mode of spoken language or, in other words, the mode of spoken language is in this instance the superior mode (as the utterance can be understood without the gesture), while the mode of gesture takes on the subordinate role (as the gesture cannot be properly understood without the utterance). Thus, we see that the modal configuration, including the distinctiveness and interdependence as well as the hierarchical structure of modes, can change quickly within an interaction.

**Multimodal methodology: means used to influence another participant**

The music lesson example also shows how one participant can influence a shift in foregrounded higher-level action in other participants by utilizing a means. In the last chapter I discussed the notion that an individual indicates a shift in foregrounded higher-level action by utilizing some pronounced deictic- or beat-type lower-level action, which I call means. Such a shift in
foregrounded higher-level action has direct influence on the person(s) that
the individual interacts with in focused interaction, as it simultaneously
forces them to adjust their foregrounded higher-level action.

The music teacher utilizes the co-construction of focused interaction
and the fact that his employment of a means has direct impact on his
student. He points to Peter in a pronounced deictic hand gesture: indi-
cating that he is shifting the foregrounded higher-level action from
practicing to verbal teaching. This deictic means leads Peter to also shift
his foregrounded higher-level action from practicing to listening to the
verbal teaching. Here, we see that a means can foster co-constructed fore-
grounding of specific higher-level actions. In a teacher–student interaction,
you will find many means that the teacher utilizes to pull the students’
focused attention from one higher-level action to the next.

**Multimodal teaching: language instruction**

A few years ago I conducted an eight-month-long study in a first-grade
classroom of a German/English bilingual elementary school. In that study
I was mainly interested in the connection between autonomy and intrinsic
motivation. Without going deeply into learning theory I would like to
revisit the example of classroom discourse from Chapter 3. There, I intro-
duced an example in which the teacher was trying to explain the word
“Graben” [“moat”] without referring to the English term. The children had
learned this word in the last lesson, and the teacher was helping the chil-
dren to remember it.

**Classroom discourse**

As you will recall from Chapter 3 (pp. 59–60), Bob is excited that he
found “a guy”. The following excerpt starts at line (16). Here, Joe and Jon
chime in, and the teacher acknowledges in line (19) that there really is a
guy that she had not focused on before.

(16) **Bob:** I JUST found a GUY!
(17) **Joe:** (pointing at own handout) a GUY right HERE.
(18) **Jon:** (gaze shift to Joe’s handout/Joe’s finger) that’s a guarder
(19) **T:** (gaze shift to Joe’s finger, then Bob) ja, da IST einer
[yes, there IS someone]
(20) **Bob:** here, right HERE (pointing at own handout)
(21) **Joe:** right there (pointing at own handout)
(22) **T:** (gaze to Bob’s handout) GENAU da ist er. [EXACTLY there he is]
(23) mhm und was MACHT der denn da oben?
[mhm and what is he DOING up there?]
(24) **Bob:** was MACHT der denn da oben?
[Bob, what is he DOING up there?]
In line (17) Joe performs a deictic gesture, exclaiming excitedly “a guy right here.” This deictic gesture does not only point, but is also a means, which can be surmised from the reactions of Jon and the teacher – who both shift their attention to Joe’s focus in line (18) and (19). From here on, the exchange is fast, without perceptible pauses, as the children are excited about the person in the handout. The teacher has refocused her attention from the moat to the person in the handout, and she develops a teaching opportunity by exploring the children’s focus of attention. Thus, the teacher shifts from focusing on her own agenda of discussing the moat to the children’s agenda of speaking about the person in the handout.

Joe’s means (his deictic gesture) indicates his own shift in foregrounded higher-level action from speaking about the moat, which he was visibly focused on in line (5), to speaking about the person in the handout in line (17). The fact that the teacher allows Joe’s means to restructure her foregrounding of higher-level actions, can be seen in line (19), where she gazes at Joe’s handout and acknowledges the existence of a person in the same line. Here, we notice that a shift in one person’s foregrounded higher-level action can (and sometimes has to) lead to a shift in a participant’s higher-level action. Clearly, the teacher could have overruled this shift in focused attention, but she uses this interest as a teaching opportunity. The teacher scaffolds and leads the students to use the target language.

When we follow Joe’s lines from the complete excerpt, we see that Joe was engaged in focused interaction with the teacher before the shift:

(5)  Joe: ground. (looking at T’s pointing finger)
(10) Joe: (gaze shift to Bob, then to Bob’s handout)
(17) Joe: (pointing at own handout) a GUY right HERE.
(21) Joe: right there (pointing at own handout)

In line (5) Joe had shared the teacher’s foregrounded higher-level action of speaking about the moat. In line (10) he looks at Bob and Bob’s handout; in line (17) Joe indicates a shift, which is taken up by Jon and the teacher;
and in line (21) Joe follows up on his previous utterance. Reviewing the complete excerpt once again, we see that Jim, Joe, and Jon show awareness of Bob and his focus on the person in the picture, by shifting their posture and their gaze towards Bob and his handout starting from line (9). When we follow Bob in the excerpt, we see that he starts out focusing on the person in the picture – not paying attention to the teacher at all:

(3) Bob: A GUY.
(8) Bob: a guy (almost inaudible)
(16) Bob: I JUST found a GUY!
(20) Bob: here, right HERE (pointing at own handout)
(25) Bob: der der [he he]
(33) Bob: die Mauer [the wall]
(36) Bob: der steht auf der Mauer [he is standing on the wall]

Bob first announces that he has found “a guy” by using emphatic stress in line (3); in line (8) he repeats his utterance almost inaudibly; but in line (16) he again uses emphatic stress to exclaim “I just found a guy,” emphasizing the words “just” and “guy”. At this point, the other children (Jim, Joe, and Jon) all shift their postures and their gaze towards Bob and his handout, showing interest, and Joe then indicates a shift in focus by employing the means following Bob’s utterance. Bob now also says “Here, right here” in line (20), and follows the utterance by pointing at his own handout.

Then the teacher speaks to Bob, and Bob starts by uttering words in the target language. From line (25) to line (36), Bob only employs the target language. Jon, who had not spoken before the shift to the person in the handout, also employs the target language:

(18) Jon: (gaze shift to Joe’s handout/Joe’s finger) that’s a guarder
(26) Jon: oh, what do you call it
(27) what do you call it
(28) tower
(31) Jon: der steht auf dem Turm [he is standing on the tower]

Of course, the target language is employed with much scaffolding by the teacher, and Jon only repeats the sentence given to him, but he does employ the target language without being explicitly asked to do so. Similarly, Bob employs the target language. But Bob has a specific interest in this person in the handout, and, when Jon claims that the person is standing on a tower, Bob disagrees in the target language by saying “Mauer” [“wall”]. Now the teacher again scaffolds and asks the question “steht der auf der Mauer?” [“is he standing on the wall?”] and Bob answers in a grammatically correct sentence, saying “der steht auf der Mauer” [“he is standing on the wall”].
Multimodal methodology: allowing a means to influence personal focus

In the above example, we find that a teacher can utilize a shift in students’ foregrounded higher-level actions as a teaching opportunity. As in other focused interaction, each participant who is engaged in mutual focused higher-level action can be directly influenced by a person’s shift in foregrounded higher-level action. This explains why the teacher shifts her focus once Joe has indicated a shift in foregrounded higher-level action, but does not shift her focus when Bob is voicing that he has found a guy.

Bob had been focused upon the person in the image from the outset; however, the teacher did not shift her attention to Bob’s focus in the beginning. When we review the first few lines, we see that Jim and Joe shared the teacher’s foregrounded higher-level action of speaking about the moat. Thus, Jim and Joe had been co-constructing the teacher’s focus, while Bob had not entered into focused interaction with the teacher before this shift in focus occurred. In lines (3) and (8), Bob’s focused higher-level action of speaking about the person in the handout was mid-grounded by the other children. Then Joe indicates his shift in foregrounded higher-level action from the teacher’s to Bob’s focus, and Jon and the teacher react to this shift. By the time that the teacher refocuses, three of her four students in the group show interest in the person in the handout.

While teachers often overrule shifts in their student’s foregrounded higher-level actions so that they can follow their own agendas, teachers can also go along with such a shift and develop the students’ interest into a teaching opportunity. Instead of pulling the students towards her foregrounded higher-level action of speaking about the moat, the teacher allowed the students to pull her towards their foregrounded higher-level action of speaking about the person in the handout.

Motivation

When the teacher refocuses her foregrounded higher-level action to the person in the handout, she validates Bob’s autonomy. Bob had been autonomous from the very beginning, but the teacher had taken no notice of him. She had only refocused once Joe, who had shared her foregrounded higher-level action, had refocused.

Through this shift in foregrounded higher-level action, the teacher is allowing Bob to be excited about the person, and to lead the group to his discovery. The teacher fosters his intrinsic motivation, resulting in Bob speaking in the target language. Jon, who had not been sharing the teacher’s focus in the beginning, is also motivated by this shift in foregrounded higher-level action, and speaks in the target language. Joe on the other hand, does not profit from this shift, even though he had initiated it.

Findings like this illustrate that teachers react to students with whom they are engaged in focused interaction, while they do not react to students...
who are autonomous. Yet it is particularly the autonomous student who profits from fostering their autonomy. As we can see in the example above, Bob takes true interest in his findings and voluntarily speaks in the target language, while Joe does not use the target language once in the excerpt.

While the teacher utilizes a shift in her students’ foregrounded higher-level action as a teaching strategy and actually taps into the students’ intrinsic motivation, we need to question this truly interactive use of means in teaching. The shift does result in fostering intrinsic motivation in Bob; however, it seems that Bob’s intrinsic motivation could have been stimulated from the outset, and we may want to start thinking about devising a teaching strategy that facilitates to foster autonomy in students who are not engaged in focused interaction with the teacher.

As van Lier (1996) points out, autonomy is closely tied to motivation and achievement, and language teaching is most effective when the teacher “stimulates intrinsic motivation, so as to take advantage of natural interests, curiosity and emergent rewards” (van Lier, 1996: 112). While many teachers are aware of this, teaching is interactively constructed and teachers are bound by communicative semantics, which facilitates shifting foregrounded higher-level actions with others in focused interaction, but does not facilitate shifting higher-level actions with others who are not engaged in focused interaction. Yet it is particularly those students who are not in focused interaction with the teacher, who are autonomous and who profit most from a shift by the teacher to their foregrounded higher-level action.

**Construction of social worlds**

| Construction of a social world: | Every higher-level action that a social actor engages in constructs the person’s social world. Higher-level actions that an individual performs simultaneously may run parallel and construct one aspect, or be disparate and construct competing aspects of a person’s social world. |

In previous chapters, I have given several examples with Sandra and Anna as participants. These examples come from a vast collection of diverse data, which I collected over a year of ethnographic study of these two women and their everyday lives. As we recall, they own a catering business together; Sandra has two children and is currently going through a divorce, while Anna has three children and is married.

In Chapter 1, I discussed the mode of proxemics, showing Anna ironing the family’s clothes while watching TV, with Katie playing behind her. There I claimed that Anna and Katie were influencing each other’s actions to some extent, due to their Anwesenheit. This claim is based on my extensive study of many different data sets, as well as my ethnographic observations.
In Chapter 4, I showed how music took on high modal density for Sandra, when the deep male voice sang “Mama, I wet my bed last night.” Again, the Anwesenheit of Sandra’s young son shapes her actions to some extent, and her son’s actions are in turn shaped by the Anwesenheit of his mother. Again, I analyzed this piece through careful triangulation – not solely based on the visual aspects of the setting. Margolis, speaking of historic photographs, rightfully warns that “it is not possible to photograph social relationships” (Margolis, 1999: 34). It is, however, possible to analyze social relationships through careful and extended ethnographic methods, including analysis of video data.

Here I use stills to demonstrate the social worlds that both Anna and Sandra are constructing: analyzed through rigorous triangulation of a vast set of different kinds of data. Since video data capture many visible and audible channels and do not record snapshots – but events – such data lend themselves to the scrutiny of social relationships and social worlds. However, as a note of caution, I would like to add here that one cannot look solely at a photograph and come to a correct analysis of the social relationships among the people in the image. Here, the image is used to depict part of the analysis, and thus is used as a transcriptual tool.

In Chapter 5, I explicated the foreground–background continuum of attention/awareness by looking at an example in which Sandra is simultaneously engaged in six higher-level actions. At this time all five children are present, while the two women are working. Simultaneously, a friend is calling Sandra, and the researcher is present. When we glimpse the women’s social world, we find that they are constantly performing several simultaneous higher-level actions: juggling the construction of their families as well as their business and their various friendships.

Shifts in foregrounded higher-level actions are often quick, and often the women’s focus gets pulled by a shift that their children perform. For example, we recall the first image in the second row of Plate 4.2, where Sandra and Anna are shopping and Anna’s younger son is attracting Anna’s focused attention by employing postures as an intense mode. Although the children had been told to stay close to the shopping cart, such requests are difficult for children to follow, and quite often the women have to pay focused attention to their children during a shopping trip.

At least some of the children are almost always present – greatly structuring Sandra’s and Anna’s days. Here, I would like to show a few other examples, illustrating how family, work, and friendships are relentlessly juggled by both women.

**Parallel higher-level actions**

The first image in Plate 7.2 shows Sandra pushing a stroller with Anna’s daughter, Katie. Sandra is walking through a shopping street to find a store in which to buy a present for a mutual friend of Anna and Sandra.
In the meantime, Anna is at home, working. This instance shows that Sandra and Anna are helping each other out to juggle their many demands.

When focusing on the image, we realize that Sandra constructs the higher-level action of shopping (by walking towards a specific store), and the higher-level action of watching the child (by pushing the stroller). The cityscape structures Sandra’s walking path, and the other people and the cars in the street influence her movement pattern, yet Sandra has backgrounded the higher-level action of walking through the city streets. During informal questioning, Sandra surmises that she takes little notice of the signs, the people, or the cars, even though she is aware of all of them to some extent when walking through the city. She further explains that she only wanted to buy that present in a specific store, and that she took Katie along so that Anna could work in peace.

Here Sandra foregrounds the higher-level action of shopping and midgrounds the higher-level action of watching Katie. Yet, the higher-level actions shift every so often. One such example is given in the second image in the first row of Plate 7.2. Here Sandra is standing in the store and cannot decide how much money to spend on the present. She is calling Anna on her cell phone, when Katie points to a picture on a box and exclaims “Käpten Blaubär” [“Captain Bluebear”], which is a German comic character. Here Sandra shifts her focused attention to Katie, and says “oh Käpten Blaubär das ist auch schön” [“oh Captain Bluebear that is nice too”], before she goes back to calling. Thus, Sandra shifts her foregrounded
higher-level action from shopping for a friend, to calling Anna, to interacting with Katie, and back to calling. While Sandra shifts her focus, the other higher-level actions are being pushed back in her attention/awareness continuum, but are all simultaneously present in her consciousness. When she pays focused attention to Katie, she is still well aware of her higher-level action of making a phone call and the higher-level action of shopping for a present.

**Divergent higher-level actions**

In another instance, Sandra and Anna are shopping for dessert dishes and glasses for an upcoming catering event.

The first image in the second row of Plate 7.2 shows Sandra, Anna, and Katie walking in a store. Anna is pushing the stroller in which Katie had been sitting, and Katie is running beside it. Their four boys are running a little further ahead of them, which is not visible in this image. Sandra and Anna are focused on the higher-level action of shopping for the dishes, and are mid-grounding the higher-level action of watching the five children. The second image of the second row shows Anna, Sandra, and Sandra’s older son David at the cash register of a grocery store. The other four children are also present. Here, Anna and Sandra, who have been shopping for one of their catering events, are getting ready to pay. David is helping, while the other children are playing at a small distance from the cash register.

Both Anna and Sandra are foregrounding the higher-level action of compiling the items in groups that will make it easier for them to package the items, transport them, and arrange them in two kitchens. Simultaneously, both mothers are mid-grounding the higher-level action of watching David: asking him to place certain items in specific spots, and are mid-to backgrounding the higher-level action of watching the four children, who are playing at some distance. While the higher-level action of shopping is backgrounded, this higher-level action certainly structures the other higher-level actions that the women perform.

**Multimodal methodology: parallel simultaneously constructed higher-level actions**

When Sandra shops for a present for a mutual friend of herself and Anna, Sandra constructs one aspect of her social world, which we can term the aspect of friendships, on various levels.

This aspect of Sandra’s social world is actively constructed: Sandra shops to buy a present for a friend, she watches the daughter of another friend (Anna), and calls Anna to discuss the present. Here we see that Sandra is actively building an aspect of her social world which not only involves the friends and a friend’s child, but also the specific store that is located in
the middle of the city that she walks to in order to buy the "right kind" of present. She largely foregrounds the higher-level action of buying the present, thereby constructing the friendship between herself and the friend that the present is for, as well as the friendship between herself and Anna. Simultaneously, Sandra largely mid-grounds the higher-level action of watching Anna’s little girl. Through this higher-level action, Sandra again constructs the aspect of the social world that we can call friendship. She has taken Katie along so that Anna can focus on her work. Finally, when Sandra calls Anna from the store to discuss the various prices and possible presents, she again constructs a piece of this aspect of her social world. When we follow Sandra a little longer, we find that she picks up the two younger boys (hers and Anna’s) from preschool, and provides all of the children with lunch as soon as the two older boys come home from school. Thus she further constructs her social world that here revolves around friendship.

As demonstrated in the example, Sandra constructs one aspect of her social world: the friendships, by constructing various higher-level actions simultaneously. As her day progresses, she continuously constructs simultaneous higher-level actions that construct this one part of her social world, and the higher-level actions run smoothly parallel to one another.

**Multimodal methodology: divergent simultaneously constructed higher-level actions**

As illustrated in the second row of Plate 7.2, Sandra and Anna are looking for glasses and buying food for a catering event.

Both women perform actions similar to those that Sandra performs in the previous example. The women largely foreground the higher-level actions of looking for dishes and placing food items on the counter, and mid-ground the higher-level action of watching the children. Simultaneously, the higher-level action of shopping is largely backgrounded. However, this time both Sandra and Anna are struggling to shift their foregrounded higher-level actions when the children demand it. The women often tell the children not to interrupt, and thereby contradict the requested shift in foregrounded higher-level actions. Also, the tone that the women employ is stricter than in many other situations that I have observed. Both Sandra and Anna try to keep their main focus on their work, and try to keep the higher-level action of watching the children in the mid-ground of their attention/awareness.

In these instances, the women perform simultaneous higher-level actions that are disparate as they construct competing aspects of their social worlds. Here, Sandra and Anna are constructing the aspect of their working worlds through buying items for their next catering event, while they are simultaneously constructing the aspect of their family worlds by watching the children. Because the two different aspects of the women’s social world
are constructed simultaneously, they compete, and you hear the women saying to each other that they do not like to take the children when shopping for their business. In conversations, they also speak about the fact that they do not mind taking the children shopping when they buy food for their families. These examples show that individuals can perform almost the same higher-level actions (shopping and watching the children) on almost the same attention/awareness levels (shopping largely in the foreground; and watching the children largely in the mid-ground) with almost the same participants (adult(s) and the children), and yet the strain for the individual juggling the various higher-level actions may be quite different.

As long as the simultaneous higher-level actions construct one aspect of an individual’s social world, these higher-level actions are easily juggled because they run synchronously. In contrast, if an individual constructs different aspects of their social world through the performance of almost the same higher-level actions, the higher-level actions may compete, and it may be more difficult for a person to juggle the various higher-level actions because they run divergently.

### Multimedia and interaction

<table>
<thead>
<tr>
<th>Multimedia and interaction:</th>
<th>People interact with others through their use of phones, computers, and faxes. By writing an email and its being read, one interacts with a person.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Real-time/timescales:</td>
<td>The concept of real-time interaction is stretched when we think about computers.</td>
</tr>
<tr>
<td></td>
<td>We write the email at our own speed (our real-time), and the other person reads it and responds according to their real-time.</td>
</tr>
<tr>
<td></td>
<td>As with “snail mail”, there is often a lag between the sent message and the response. When we think about instant messaging, the response time is shortened and comes much closer to what we originally called face-to-face interaction.</td>
</tr>
<tr>
<td></td>
<td>When thinking of interaction as happening on different timescales, we no longer have to differentiate between face-to-face interaction and computer/phone-mediated interactions on the grounds of time.</td>
</tr>
<tr>
<td></td>
<td>Real-time is not one time, but rather is made up of various timescales.</td>
</tr>
</tbody>
</table>
Whalen *et al.* studied the work practices of a sales representative in an office equipment company and found that these practices were “organized through an improvisational choreography of action involving not only the turn-by-turn interchange with customers on the telephone but also the concurrent utilization of a variety of tools and artifacts” (Whalen *et al.*, 2001: i).

Such choreography of action involving diverse tools and artifacts can be found in many office settings. While Whalen *et al.* focus on one person choreographing his actions in accordance with specific telephone calls, let us now return to the accountant’s office.

As shown in earlier chapters, the accountant interacts with his assistant in her office, utilizing many modes: including the disembodied mode of print in the form of a document. When we follow the accountant into his office, we see how he interacts with various people simultaneously by utilizing a variety of tools and artifacts.

The transcript in Plate 7.3 precedes and follows the interaction between the accountant and his assistant that I discussed in the last chapter. There, the accountant had utilized the three audible means to indicate a shift in his foregrounded higher-level action. We saw that the assistant invoked a delay in his actual shift, but, after a brief exchange, the accountant does complete the shift and walks over to his office.

The first four images in Plate 7.3 illustrate the accountant’s work practice before the previously discussed interaction with his assistant, and the next four images illustrate his work practices following the interaction. As mentioned in the last chapter, the accountant has a visitor, with whom he continuously interacts while in his office.

The first two images in Plate 7.3 show the assistant standing in the door of the accountant’s office announcing a call, and then the accountant taking the call. The next image in this row shows the accountant conversing with his visitor. The first image of the second row illustrates that the accountant is holding a computer mouse in his right hand and looking at his computer screen. The second image of this row shows that he has turned around slightly, opening up his posture – he is now gazing directly at the visitor and they are engaged in conversation.

Between these four transcribed images and the next images in this plate, the interaction between the accountant and his assistant takes place in her office. As soon as the accountant returns to his office, in which the visitor is still sitting in the same spot, he sits down at his desk and signs the document, as illustrated in the first image of the third row in Plate 7.3. Following the signing of the document, the accountant stands up and is again involved in conversation – gazing at his visitor as illustrated in the second image of the third row in Plate 7.3. Here, the accountant is holding the signed document in his hands, while speaking about a different topic with the visitor.

In the first image of the last row, we see that the assistant has come into the accountant’s office, and the accountant is handing her the document. No words are spoken between the accountant and his assistant at
Plate 7.3 Transcript: multimedia and interaction.
this time, and the accountant converses without perceptible pause with his visitor. The last image illustrates that the accountant has taken a seat at his desk again. He is still involved with his visitor and has again taken hold of the computer mouse.

These snapshots of the accountant’s work practices show that he is choreographing several personal interactions – the interaction with his assistant, the interaction with the visitor, and the interaction with the client on the phone – while he is at appropriate times utilizing the computer and signing a printed document. This very short, about seven-minute-long, excerpt of one of his days illustrates the quick successions and sometimes overlapping actions that he performs.

All of these choreographed actions take place in real-time. But what are the different scales of real-time here?

Lemke suggests that:

each act participates in local constructions of meaning on shorter timescales at the same time that it also participates in the systematic networks of interdependent activities that sustain institutions and societies over much larger distances and longer times.

Lemke, http://www-personal.umich.edu/~jaylemke/ (accessed 25/02/04)

This historical view of the extent of actions, simultaneously constructing local meaning as well as institutional and social meaning, is important for our understanding of human interaction on a micro and a macro level.

In addition, the notion of time needs to be considered for a multimodal analysis of interaction on different levels. Real-time is not just one concept, but rather is a multitude of concepts that are categorized as one. While some actions fall at one and the same time, so that the accountant, for example, signs the document while the visitor is sitting in his office, or the assistant is picking up the form while the accountant is speaking with the visitor, all of these actions started at a different moment in time. Every action has a different duration and will likely also be completed at a different time.

The concept of a modal density foreground–background continuum can give us insight into the workings of human interaction beyond focused interaction. When we analyze the higher-level actions that the accountant engages in simultaneously, we discover that he is constantly engaged on some level with his assistant (the foreground, mid-ground, or background). His assistant can enter his office when he is engaged in focused higher-level action with the visitor, and the accountant interacts with her, giving her the floor when needed (i.e. when she needs to give him a message) or by simply handing her the document. Simultaneously, he is engaged in interaction with the visitor – at times foregrounding the higher-level action of conversing with the visitor – and mid- or backgrounding this
action if other higher-level actions are more important at a certain moment in time. When he speaks on the phone, he gives the client his focused attention and, when reading something on his computer, his attention is briefly focused on the screen. Similarly, when he signs the document his attention is briefly focused on the document.

While there is certainly much simultaneity in higher-level actions, we need to remember that the starting points, the duration, and the completion of the higher-level actions are usually quite different. Also, when, for example, tracing the higher-level actions that involved the document, we will see that the assistant is engaged in a higher-level action: filling out the document, and the accountant is also engaged in a higher-level action: signing the document. These higher-level actions are somewhat overlapping (when the accountant is answering the assistant’s questions about the form) and somewhat distinct (when the assistant fills it out and when the accountant signs it).

Depending upon the focus of study, we can heuristically tease out the actions that each participant is performing, and thereby focus on particular social actors, or we can look at the actions that are structured around the document, and thereby focus on the document. However, we focus, we can trace the higher-level action on a timeline within the office, and we can consider the historical timeline of the document and also the social and organizational meaning of it. The actions that the accountant and the assistant perform in less than seven minutes concerning this document are only a part of the sequence of actions that evolve around it.

While the higher-level actions that the accountant and the assistant engage in involving the document are fairly short-lived, there are other actions that are occurring simultaneously, which have a much longer duration.

The higher-level action of interacting between accountant and assistant, for example, has been occurring for many years. There is a routine, a pattern, and knowledge between these two social actors that has evolved through a history of encounters. The assistant has asked many questions and the accountant has signed many documents. The assistant has entered the accountant’s office many times when he has had visitors, to announce an important call. Thus, while these particular higher-level actions are occurring at this time, both social actors have acquired what Nishida (1958) calls a historical body, or an understanding of each other’s actions and ways of being, that make the interaction run smoothly.

Just as the accountant and the assistant have developed mental representations and processes of discourse among themselves, they have also developed mental representations and processes about the use of documents. The timelines of the higher-level actions that involve the document and involve the interaction between accountant and assistant are, however, of different durations, different structures, and different types. Yet, one impacts the other.
Thus, the accountant not only choreographs the higher-level actions that involve the turn-by-turn interchanges with his assistant, his visitor, and his client on the phone (by utilizing a variety of tools and artifacts), he also juggles the different timescales that each one of these higher-level actions entails.

**Assignment**

Carefully look at the example in the accounting firm. Pick one particular instance and draw modal density circles for the higher-level actions that the accountant performs at that moment. Then place the various higher-level actions on a foreground–background continuum. Now look once more at the transcripts (Plates 6.3 and 7.3) and note down the levels of attention/awareness that the accountant places on each higher-level action at each point illustrated.

Videotape an interaction and analyze some of the complexity that you find. Share your video and your analysis.