

Sjoerd Ketelaars, BA

s4591267

Master Thesis Linguistics

Dr. J. G. Geenen & Dr. O. N. J. C. Koeneman

3 February 2020

Turning the Tables? An Analysis of Turn-Taking within Conversation

Sjoerd Ketelaars, BA – Radboud Universiteit Nijmegen – MA Language and Communication

Coaching

s.t.ketelaars@student.ru.nl

Table of Contents

Abstract	p. 3
1. Introduction	p. 4
2. Literature Review	p. 7
2.1 Multimodal Discourse Analysis	p. 7
2.2 Conversation Analysis and Turn-Taking	p. 10
2.3 A different Approach	p. 16
2.4 Two separate systems	p. 21
3. Empirical Methodology	p. 24
3.1 Task and Protocol	p. 24
3.2 The game	p. 26
3.3 Participants	p. 27
3.4 Recording	p. 27
4. Analytical Methodology	p. 29
4.1 The unit of analysis: Mediated actions	p. 30
4.2 Lower and higher-level actions	p. 32
4.3 Modal configurations	p. 34
4.4 Modal density: the foreground-background continuum	p. 34
4.5 Conclusion	p. 36
5. Analysis	p. 38

5.1 Figure 1 – Adaptation of Production Length (APL)	p. 38
5.2 Figure 2 - Shift in Mode Use (SMU)	p. 47
5.3 Figure 2 – Cause of Cross Boundary Effects (CCBE)	p. 52
5.4 Figure 3 – Additional evidence	p. 56
6. Conclusion	p. 66
7. Discussion	p. 69
7.1 Implications for the turn-taking model	p. 69
7.2 Online Production Processing	p. 72
7.3 High efficiency approach	p. 74
7.4 Future Research	p. 75
References	p. 78
Appendix A: Script Instructional video	p. 81
Appendix B: Essay Cover Sheet	p. 82

Abstract

Turn-taking as a concept has long had its influence within the field of Conversation Analysis. Since it was first introduced by Sacks, Schegloff and Jefferson (1974), it has been analysed and refined into the field of modern psycholinguistics (Levinson, 2016). The current thesis looks both at theoretical evidence and data from a naturalistic, explanatory, two-party task. Both of these show there are certain problems within the turn-taking system as it currently stands. Not only does it come short in properly capturing the multimodal nature of communication, it also is unable to explain certain constant and ongoing communicative exchanges within classical turn boundaries.

This thesis will argue that a less rigid system in which turns are considered to be mainly temporal units is better suited to explain the realities of day to day communication. Communication should be considered more fluent and constant, with so-called Inter Boundary Modulations (IBM) being a prevalent and very normal phenomenon, in which traditional 'listeners' are able to influence the online decision making of the traditional 'speaker.' The analysis in this thesis will put forward three of the most common types of IBM it has been able to reveal within the current analysis. In sum, the conclusion will be that the rigidity of the turn-taking system must be revised as well as its psycholinguistic interpretation, to leave more room for the influence of online interpretation on online production.

Keywords: turn-taking, Inter Boundary Modulation, multimodality, online interpretation, online production, Cross Boundary Effect

1. Introduction

The way humans interact is something that can arguably set them apart from other animals. Human communication functions at incredible speeds and with unprecedented levels of complexity. It is therefore not surprising that modern research has found a place for fields such as linguistics and, more recently, multimodal interaction analysis to attempt to dissect these processes in more detail and discover how humans navigate the complexities of interaction and communication. To further the search for an answer to this immense question, the current research will focus on the organisation of everyday interaction and communication.

Regardless of language, cultural background or heritage, all forms of interaction require a form of organisational structure in order to allow for an efficient flow of information. One existing and fairly widely accepted system to do this is that of turn-taking, which structures language in a roughly linear pattern with interlocutors speaking mostly one person at a time and moments of overlapping communication being common but brief. The current research will challenge this approach by instead proposing a different kind of system. This thesis will argue that when people are interacting, there is a lack of distinct boundaries in turn-taking behaviour in relation to propositional content. While producing, there is in fact simultaneous interpretation and this interpretive process is not only feeding online speech production, but is an integral part of it.

One of arguably the most iconic ways of identifying language is that of the conduit metaphor, as originally presented by Reddy (1979). He describes communication as a conduit for the transfer of ideas, feelings and concepts from one person to another (Grady, 1998, p. 2). Within this model, one can use language as a conduit to encode ideas and then send them to a willing recipient. The current research challenges this assumption by opposing it by an idea formulated by Schefflin (1964), stating ‘the intent of an interactant and the function that a behaviour actually has in a group process must be conceptually distinguished’ (p. 318). Any

behaviour can be communicative, whether the intent is there or not. This means language cannot function as a pure conduit for the transfer of ideas, since it is not a process undertaken by a single party. This thesis will argue that people are aware of the fact that their communicative processes are in large parts dependent on their fellow interlocutors and are therefore constantly tapping into communicative actions as produced by their listeners. This helps them make the best possible online decisions in order to allow for communication at the highest possible level of efficiency. The fact that these two processes of interpretation and production are so closely related might also be a part of the reason why people are capable of switching roles so quickly: their planning and production coincides with their interpretation, rather than the two competing for cognitive resources.

To establish this claim, the current research will present historical and recent ideas on communication and specifically on turn-taking. Turn-taking is one of the organisational systems that has been devised within conversation analysis as a more or less universal organisational model for interaction. The current research will present some new and existing criticisms to this model as it was first outlined by Sacks, Schegloff and Jefferson (1974), that suggest it occasionally contrasts with reality to a level that makes it difficult to uphold. The most relevant objections for the current proposal are that it is said to not incorporate modes of communication other than spoken language (Power and Martello, 1986) and that the incredible pace at which turn-switching takes place seems difficult to match with current psycholinguistic theory (Levinson, 2016). The way turn-taking and more specifically turn-switching is explained does not match with the current research on human cognitive ability, as well as the explanation relying heavily on the notion of completeness that once again seems to not directly match with communication as it is regularly observed in daily naturalistic contexts.

The current research has analysed naturalistic data from a total of 14 Dutch university students in an attempt to find instances where the turn-taking system as presented is undermined. It will present three examples in which this system is seemingly not working as presented and discuss possible reasons for this. These examples are situations in which it is made visible how the online communicative processes of an interlocutor are influenced by received input from a listener, so called 'Inter Boundary Modulations', or IBM for short. IBM's are the crucial moments this thesis will argue problematize the concept of turn-taking. They are the specific moments in time in which interaction between what classically constitutes a speaker and a listener unfolds within classical turn-boundaries, because a speaker interprets and responds to communicative actions as undertaken by the listener in real time. Sometimes speakers will even interpret certain actions that only influence their communicative choices later on in the interaction, a situation called a Cross Boundary Effect, or CBE. This thesis will look towards work by Goodwin (1980) and recent work by Pirini (2016, 2017) to suggest a different kind of model where communication is instead a non-linear flow of intertwining cognitive processes. In this model, all social actors of any communicative process are aware of the fact that they all contribute to the meaning and flow of said process and will thus all attempt to have this run as smoothly as possible. Speakers are not only attempting to send a message, they are attempting to send a message to their targeted listener or listeners and will use the input of said listeners to do so. All involved parties are therefor in a constant process of information exchange and all propositional content is malleable to the highest degree. People effectively work together to inform one another on the progress of their actions and make a collective effort for this communicative process to have the best possible chance at success.

2. Literature Review

Multimodal Discourse Analysis has risen as a prominent field of research within the organisation of human interaction. Spoken language was previously considered the main channel of communication, but is now recognized as only one of many modes of interaction (Schefflin, 1964, Goodwin, 1980, Kress and van Leeuwen, 2001, Norris, 2004, 2009, 2011, 2013, Pirini, 2016, 2017). This has had significant consequences for the conceptualisation of conversation organisation. Conversation analysts originally built a framework focussing on interlocutors' moment to moment turn-taking (see f.e. Sacks, Schegloff and Jefferson, 1974) and the sequential development of these turns. However, more contemporary evidence suggests that this type of sequentiality is perhaps not the correct way to view the organisational structure of interaction. Analysis as performed in this thesis is able to reveal instances of so-called Cross Boundary Effects and Inter Boundary Modulations, which are elements that could not fit into a system with a strict division into sequential turns. This chapter will provide some of the background and origins of the field of Multimodal Discourse Analysis and its relation towards social semiotics and conversation analysis, while also highlighting some of the problems that have risen due to more contemporary additions within this body of literature. Finally, it will turn to the previously mentioned Cross Boundary Effects and Inter Boundary Modulations as elements that demonstrate the real-time problems a rigidly sequential turn-taking system has.

2.1 Multimodal Discourse Analysis

Multimodal Discourse analysis originated from two major shifts as they occurred within the preceding field of discourse analysis. The first was initialized by Kress and Van Leeuwen (2001) as they shifted the focus of discourse analysis from being a study of language, to a study of modes. They built this idea by taking social semiotics as posed by Halliday (1978) to organise the structure of all these communicative modes. Within social semiotics, these

modes are importantly conceptualised as being socioculturally shaped and constructed. This makes them both highly dynamic and contextual. The challenge of a social semiotician, is therefore to articulate the complexities of human meaning-making behaviour by attempting to elucidate the ways in which these ever-changing systems are used and manipulated (Geenen, Norris and Makboon, 2015). Social semiotic multimodal discourse analysis thus tries to determine the link between the use of any semiotic resource and their functional aspects (Van Leeuwen, 2005, as qtd. in Geenen et al., 2015). These resources can be anything from objects such as clothing or computer software, to language and posture. Within social semiotics, a strong focus is put on the fact that all these modes are fluid and open. They are constructed by meaning potentials that have become embedded within them through past uses and developments over time. A social actor can then convey meaning by making choices across these different semiotic resources, accessing the ones they see fit based on each system's affordances and limitations they have received over time. Because where there is choice, there is meaning. From a multimodal perspective, meaning is also closely linked to how these different semiotic resources interact with each other. The main addition of social semiotics is therefore that it has expanded the scope of discourse analysis by incorporating non-linguistic communicative modes into the multimodal analysis.

The second major shift finds its origins with the change from considering language as the organizing system of human interaction to action being considered the organizing system of interaction (Scollon, 2001; Vygotsky, 1978; Wertsch, 1998, all qtd. in Geenen et al., 2015). This change effectively makes it so that the social actor is now incorporated within the system, as the action is in fact the social actor acting through mediational means.¹ These ideas have their roots in anti-reductionism as it was originally championed in sociocultural psychological theory, most famously by Vygotsky and his contemporaries. The most relevant

¹ A more expansive overview of the mediated action can be found in Chapter 4, Analytical Methodology

contribution from their perspective lies with the notion of mediation. Vygotsky stated that all human (inter)action is in fact mediated. In his own work he focussed on the ways language mediated learning and development, but its conception can extend to all forms of social action. Vygotsky also argued for an antireductionist methodological approach, as he saw that isolating the component parts of a phenomenon to help with analysis can often be misleading. Wertsch (1998, as qtd. in Geenen et al., 2015) picked up on these ideas and pushed the mediated action to the front even more, considering them as the ideal unit of analysis since it forces an analyst to always consider both the social actor and the mediational means. Scollon (1998) took a similar stance, using the mediated action as a way to guarantee the consideration of the complex and irreducible tension between social actor and mediational means. Scollon (1998) stipulated three important principles as the theoretical framework for multimodal discourse analysis: (1) social action, stating that discourse is best understood as social action; (2) communication, stating that an action can only be social when meaning is communicated; and (3) history, stating that all social actions are affected by historical elements (Geenen et al., 2015).

With these changes incorporated, discourse analysis was able to develop into multimodal discourse analysis. The main two differences are the fact that language is no longer considered the a priori leading mode, meaning all other modes can potentially relay similar or higher levels of communicate information, as well as the mediated action now being considered the most important unit of analysis. This last change allowed for the importance of the interrelationship between social actors and their mediational means to fully come through in an analysis. These changes also had their importance on other more specific areas of discourse analysis, as for example the next section will show.

2.2 *Conversation Analysis and Turn-Taking*

Conversation analysis as a discipline originally mainly focussed on naturally produced speech produced by interlocutors in various everyday settings. A big focus of this field was that of turn-taking, the sequential ordering of the moment-to-moment contributions as produced by participants. A lot of work towards this phenomenon was done by Sacks, Schegloff and Jefferson (1974), who tried to find order in the chaos that is conversation by describing the organisational structure of a turn-taking sequence. They state how speakers are said to very finely order the way they speak. Participants primarily talk one speaker at a time and while the length and order of participants may vary, transitions between speakers are neatly coordinated. Sacks, Schegloff and Jefferson took this as their central proposition about the configuration of human conversation and compared it to a large set of empirical data, in order to analyse the phenomenological components of natural speech and conversation. In total, they analysed fourteen elements they claim exist in any conversation and that therefor need to be accounted for in order for any sort of organisational system to work. The arguably most defining characteristic of this system is that it is meant to deal with single transitions at a time. It aims towards one person speaking at any moment who then shifts the turn towards another participant or simply ends their own turn. The first three of the fourteen elements clearly show this outlook:

- (1) (1) Speaker-change recurs, or at least occurs
- (2) Overwhelmingly, one party talks at a time
- (3) Occurrences of more than one speaker at a time are common, but brief

Sacks, Schegloff and Jefferson, 1974, p. 700.

They continue their argument by showing how turns can often clearly be observed, either via direct and overt turn-allocation techniques, as well as the occurrence of repair mechanisms to fix problems such as turn-taking errors or simultaneous speech. The

construction of the turn themselves are subject to a concise set of rules. A person will start a turn if either (1a) the turn is assigned to them, (1b) they take the turn by speaking first after another turn has ended or (1c) the speaker starts another turn after their initial turn ended and no one else stepped in. If (1c) occurs and the speaker takes another turn, step (1a) to (1c) reapply at the end of that turn (2) (p. 704).

This model of turn-taking in interaction has proven incredibly influential over the years that followed. It is generally considered to be the most universally stable communicative system (Levinson and Torreira, 2015), applying to practically all languages with only small deviations. It matches up with most well-known and common auditory phenomena, with conversation almost always ‘sounding’ as if it is spoken in a continuous ordering going back and forth between speakers. Certain types of speech have been raised to break with this pattern, such as lectures and press conferences, but these are often culturally motivated rather than universals (Levinson and Torreira, 2015). It does seem like this organisational form is in fact the default way of communication, as it is what people revert back to when talking to friends and family or for example in the context of language learning. There has been quite a lot of work done to add to the understanding of the abovementioned rules (see Clayman, 2013; Drew, 2013; Hayashi, 2013 for overviews, as qtd. in Levinson and Torreira, 2015), but there are also certain problems that Conversation Analysis runs into when considering turn-taking. The first problem is the question of what exactly constitutes a turn. There are occasions when a turn can be no more than a single word, while other times they consist of complete stories. This question receives another layer of difficulty when one considers the multimodal nature of conversation, because at what point would a certain gesture, head movement or posture shift be communicative ‘enough’ to be considered its own turn. A second large issue that turn-taking faces is the incredible speed at which people are able to switch turns. Time between turns tends to be only around 200 ms, which is roughly the

length of a single syllable. This therefore seems to be at the limit of human performance (Levinson, 2016, p. 7). Within conversation analysis, it is thus generally believed that there is extensive prediction during comprehension, since interlocutors have to plan their response during the turn of their conversational partner. Effectively, a listener must thus not only comprehend what a speaker is saying, but also predict how the speaker is going to finish his turn and when he will yield his turn. This seems to be the only possible way, since the known speed for conceptualisation is already equal to the 200 ms between turns, while the speed for actual form encoding takes nearly twice as long in addition to that. A listener will then start preparing his own response to make sure he is ready to launch his own response as soon as the current speaker is finished.

Recent work by Levinson (2016, and Torreira 2015) provides some evidence that this is indeed the way conversations unfold. One such piece of evidence comes in the form of EEG studies that can show certain N400 effects before a sentence is finished. The N400 is a type of observable brainwave that is observed roughly 400ms after certain stimuli. Its amplitude is most notably affected by semantically or orthographically deviant words. Levinson and Torreira mention the example (2015, p. 18) of the sentence ‘she carried eggs in a ...’ in Spanish. This sentence would then be followed by either ‘a basket (*una canasta*)’ or ‘a sack (*un costal*)’. Because of the differently gendered pronouns in Spanish, the N400 effect was already observed before the actual noun was encountered. According to Levinson, it is effects like these that show the level of prediction humans undertake while listening in order to correctly formulate their own responses in time. A first objection to this however, is the fact that research into the N400 effect is highly controlled. An overview article by Kutas (et al., 2011) shows how effectively all these effects are measured within controlled and organised experiments. This is not unexpected, since current EEG and other types of neurophysiological research is highly constraint by the recording setups. The recorded N400

effects are therefore almost universally acquired from situations with only very marginal ecological validity. It is therefore difficult to directly link this to the fact that humans devote extensive resources to predicting their interlocutor's next utterance. Natural conversation will always have a theme or a topic. Interlocutors will have already developed certain expectations and knowledge about each other as well as the conversation that influences their expectations concerning the interaction that is taking place. These are all factors that are not, and perhaps at this point cannot, be taken into account in this type of research. The unpredictability that is an intrinsic part of natural language can at this moment not be claimed to be fully accounted for considering these N400 effects as they are currently presented.

A second element that these experiments are unable to capture fully, is interaction's multimodal character. Apart from associations made by interlocutors about each other and the flow and topic of a conversation, natural conversation never purely consists of speech. The input a listener receives through gestures, posture shifts or facial expressions can heavily influence what is considered 'unexpected'. In isolation, one might not expect a sentence like 'The cake looked gorgeous,' to be followed by 'and tasted horribly.' This completely changes if the speaker already has a disgusted look on their face during this first part of speech. Now the fact that there was some form of twist coming to this statement might be exactly as expected. By taking into consideration all the modes that are employed by an interlocutor, the expectations and potential predictions of the interaction have been completely changed. A listener is able to tap into all modes of communication as they are presented to them, which can change both meanings and expectations. It is also relevant to keep in mind that the fact that a listener has to continue to interpret all these different modes is also another burden towards a listener's cognitive load. The following paragraph will zoom in on this topic more, but it is relevant to keep in mind that a 'listener' is never only listening, but also watching and even smelling or feeling.

Within the amount of cognitive load lies another predicament concerning turn-taking, one that Levinson himself admits to. The system he presents would require a level of multitasking that is considered notoriously difficult in light of current psycholinguistic theory (2016, p. 9). His model would underlyingly suggest that there is effectively only full comprehension during the initial part of an utterance, while as soon as a listener believes he can sufficiently predict the outcome of the utterance, he or she starts diverting cognitive resources towards the planning of his or her own production instead. His theory suggests that a large amount of planning is required to explain the incredible speed at which interlocutors are able to switch turns. Listeners therefore predict the endings of turns so they can start the planning process early, in favour of listening to the full sentence. This could effectively have two consequences. Potentially, it could mean that input received in the latter half of a sentence is sometimes just plainly ignored. The brain is engaged in the planning task, therefore has to drop the comprehension of this input. The other option is that the brain is capable of both planning and comprehension at the same time, but it would require serious re-planning of any form of prepared response if the latter half of the sentence contained any unexpected information, resulting in slower reaction times. Levinson is able to provide some evidence for the second option (Bögels et al. 2016), but this immediately also presents him with another problem. If at this point a listener is still listening and receiving all elements of the speaker's input, that means from the moment they start planning they are now simultaneously predicting, listening and interpreting as well as planning their own response. And as mentioned in the previous paragraph, they are interpreting not only speech, but also posture, gaze, object handling and many other modes involved at any point in the conversation. At this point one could also reasonably raise the question what the point of planning even is, since listeners do still seem to be ready to change or deviate from previous predictions if the situation calls for it. If they are planning, one would have to assume they are planning very

broadly and are perhaps ready for multiple potential outcomes. All of this seems to only ever increase the cognitive load a person is apparently capable of dealing with, to the point where it is perhaps also looking less and less likely that this is the complete story.

A final problem then, lies within the notion of completeness. The way this model is currently framed suggests that there are three things a listener does while still listening. In Levinson's own words (2016, p. 8): A speaker will start 'Speech act prediction', in which their 'response planning begins'. Then they will switch to 'Turn-end prediction', in which they start to predict the end of the current turn. And finally, they will wait for actual 'Turn-ending cues', which they will use as their own 'production launch signal' to start speaking themselves. As mentioned before, Levinson assumes here that because there is simply not enough time to do all of these things between turns, a listener will start these processes while still listening to the current speaker in order to be prepared for when they have to switch into the speaker role themselves. This assumption seems problematic however, as Levinson assumes that all of these processes need to be completed in full before someone can start speaking. Within this framework, a speaker needs to have their speech act fully planned out before one is able to produce it. This is something that does not seem to match actual real-life conversations. It is very rare for a conversation to consist of nothing but complete and well-structured sentences. This then seems unexpected, because raises the question why there would be so many incomplete or incorrect sentences and even incorrect words if every person always devotes considerable cognitive resources to planning out and preparing every utterance. Normal conversation regularly contains syntactical anomalies and half-sentences. Sometimes people will even drop certain propositions halfway through, realizing that their thought processes were flawed or the words they are using do not best fit what they are trying to convey. This seems to be at odds with a system where every person would plan out an utterance in full before speaking it, especially if they were prepared to sacrifice some level of

attention towards the speaker that they are responding to. Taking all of these considerations together, it seems that a model of prediction and planning carries a lot of unanswered questions that make it more difficult to match with reality than may have appeared at first glance. The following section will therefore attempt to take a different approach to the organisation of conversation, employing a more open structure with a more constant back and forth flow of communication.

2.3 A different approach

It appears plausible that an alternative model may be more accurate in representing the psycholinguistic realities of natural moment-to-moment conversation. Within the turn-taking structure as it has been devised over the years, it may be the case that auditory phenomenon have perhaps been inappropriately conflated with the cognitive function underlying those material realities. This thesis will therefore argue that communication is a constant process of both production and interpretation, which results in a lack of clear turn boundaries. These two processes are not only intricately linked, but interpretation is in fact an integral part of speech production. There is a constant sharing and exchange of information through Inter Boundary Modulations and Cross Boundary Effects, which allows for a more highly efficient transfer of information as all resources constantly tap into each other. In order to use this to its maximum potential, propositional content is continuously malleable to a very high degree and can be adapted on the go. This section will expand on these effects and show the theoretical motivation for their existence, as well as how they are difficult to match with existing turn-taking structures.

It is arguably not possible for there to be communication without communicative intent (Recanati, 1986). In this sense, whenever people communicate with one another this would be the first thing to look for. Is there communicative intent coming from a potential interlocutor and if so, what is it? The answer a person gives themselves on such a question

effectively decides whether communication is going to take place or not. Considering the importance of this answer, people are likely to be constantly vigilant for such intent, making sure that all actions at any moment are interpreted and adhered to. However, this need for communicative intent should be examined with care and precision. As Schefflen (1964) put it: 'In the first place, human behaviour can be communicative, whether or not it is intended to communicate' (p. 318). At surface level, this might suggest that this invalidates the earlier claims. Communication does not require the intention to communicate at all, unless one were to narrowly interpret communication as akin to the conduit metaphor (Reddy, 1979). Within this metaphor, communication can only occur through the active use of words as a projector of ideas, which another interlocutor can then tap into. Schefflen however then adds to his comment by stating that 'the intent of an interactant and the function that a behaviour actually has in a group process must be conceptually distinguished' (p. 318). This is what brings us to the crux of communicative intent. Behaviour by any interlocutor has the potential to be analysed as having a function within a group process, or in other words, as communication. If two people are sitting down together and one person wipes their forehead this could have no form of communicative intent from their perspective, with the person perhaps simply having an itch. If the other person interprets it as such, now the 'wiper' has done nothing but unintentionally relay the fact that he has an itch. However, if a person interprets it as a way for the 'wiper' of relaying they are feeling warm, they might then respond to it by asking if they should open a window. Suddenly, this simple gesture to get rid of an itch has been turned into a communicative action which then sparked more communication in the form of a question being posed. The only thing that has initiated this change was the fact that the act of removing an itch was interpreted as an action with a certain communicative intent that was perhaps never there from the 'wiper's' point of view.

The situation described here has some interesting implications. In this situation an action was performed with no intent of being communicative and in no way meant to invoke a response. However, the interpretation of the listener turned it into the start of a conversation. If a third person party observed this interaction, they would probably now describe the wiping of the forehead as the communicative that directly sparked the conversation that followed. But how could this possibly have happened if the 'wiper' never meant it this way? The answer is implicitly there: because the listener interpreted it as such. As described by Scheflen, the intent by the 'wiper' is separate from the function the behaviour has to the listener. The situation now illustrates a scenario in which the listener has directly influenced and arguably even changed the 'communication' produced by the speaker. One could now consider the possibility that people are, perhaps unconsciously, aware of the fact that listeners might attribute their own meanings and intentions to any communicative action. This would break up the conduit metaphor as it inserts an area of uncertainty where the initially encoded message might not be what is received in the end. Therefore, speakers would presumably attempt to make use of all available resources to the best of their ability to encode their propositions in a way that gives the best chance of their target audience interpreting things in the way they want them to. And arguably, the target audience itself would be the best resource available to do so.

Work by Goodwin (1980) already made some suggestions towards a type of constant adaptation. He described it as 'In conversation speakers are thus faced not simply with the task of constructing sentences but also with the task of constructing sentences for hearers' (p. 277). People do not simply decide on their utterances by their own intent, but also by what they notice and interpret from their interlocutors. In order to do this most effectively, a more flexible approach to the structure of turns seems necessary. Rather than each party sequentially interpreting, predicting, planning and responding, there is room for so-called

'Inter Boundary Modulations.' This concept ties in with the problems concerning the notion of completeness as described in section 2.2, as it refers to small instances of communicative actions undertaken by the interlocutor within a conversation who would normally be considered to be the listener. Importantly, they are the party who are currently not at turn to talk, but are still being interpreted by the current speaker who makes modifications to their modal configuration because of the actions they interpret. This immediately brings up the second modification needed to the turn-taking system: propositions are not planned out in full. As a matter of fact, they are malleable and adaptable to the very highest degree and to the last possible moment. This allows a speaker to adhere to all potential input they receive during their turn at-talk in the form of these Inter Boundary Modulations. Goodwin provides a simple but important example for such an occurrence:

(2) Suppose that a recipient begins to display proper hearership well after the speaker has begun to produce a sentence. If the speaker brings that sentence to completion, the utterance will contain a coherent sentence and no sentence fragment. However, when the actions of both speaker and hearer are taken into consideration, that complete sentence may in fact constitute a fragment since only part of it has been attended to properly by a hearer: By beginning a new sentence when the gaze of the recipient is obtained, the speaker is able to produce an entire sentence while being gazed at by the hearer.

(Goodwin, 1980, p. 277)

This situation describes a very strong example of the communicative system that is argued for in the current research. The reaching of mutual gaze actively changes the approach taken by the speaker, with them deciding to restart their sentence. However, the listener might in fact have been listening from the very beginning. The interpretation of the speaker that this reaching of mutual gaze meant that the listener 'started listening' has triggered them to restart the sentence. This gaze shift by the listener can therefore now be described as an Inter

Boundary Modulation. It was their action that sparked the speaker to adjust their utterance into an utterance that now contains a fragment sentence. What this seems to show, is that the speaker was not only speaking, but already actively looking for a connection and communicative signals they could receive and interpret. More importantly, the speaker was looking for signals to adapt to and was prepared for productive communicative actions originating from the listener. It could of course equally be the case that the speaker came to the right conclusion and that indeed the listener was not listening from the start. This would potentially mean that the listener consciously signalled their attention by gazing at the speaker. In that case, the listener made an effort to relay the fact that they were previously not yet ready to listen and this sparked a change in modal configuration by the speaker. Inter Boundary Modulations can thus both be conscious or unconscious, just like any other communicative action.

From the existence of these Inter Boundary Modulations almost naturally also flows the existence of so-called 'Cross Boundary Effects'. If speakers adhere to all communicative actions of listeners as they undertake them during their own turn, they will presumably use this information during the whole of their conversation, including any future turns. These moments are called Cross Boundary Effects, and they refer to times where speakers make choices based on information they have gathered throughout the earlier parts of the conversation. This can be for example expecting a certain response they received once before or employing a certain mode in an attempt to meet a listener's preferences. These effects are again testament to the fact that there is constant interpretation during production. They also show how highly this information is valued by speakers, as they not only influence the speaker directly during their turn but even across turns and into later parts of the conversation.

Some suggestion towards this type of continuous-exchange approach has already been made by Pirini (2016). He explains how certain higher level actions, such as tutoring, are in

fact co-produced. The presence of another party affects the approaches to the action of both of the interlocutors, even though they are essentially still both producing their own higher-level actions. While it may seem logical that these types of actions are always influenced by multiple interlocutors, it shows initial evidence of mediated actions being influenced and produced by multiple participants. If human beings are capable of this type of co-production, it does not seem unlikely that they apply this in more day to day conversations as well. In later research by Pirini, he also concluded that ‘to treat interaction as always sequentially produced and sequentially contingent misses the global nature of higher-level actions’ (Pirini 2017, p. 125). He analyses co-production in terms of Norris’ (2004) concept of modal density on the one hand and attention levels on the other hand.² Pirini’s analysis of co-produced higher level actions shows how every participant involved in the co-production assigns different levels of modal density and attention to each higher-level action they perform. This means that they each perform their own higher-level action, which together co-produces one big higher-level action. This does not happen sequentially, both participants adding to it one after another, but instead all happens in one go.

2.4 Two separate systems

The two systems have now been laid out and as mentioned before, they do not appear directly compatible. The classical turn-taking system as presented is ordered in a fairly rigid sequential pattern. Turns follow each other in a continuous line, each turn being followed by the speaker taking another turn or another speaker starting their turn. There does exist some criticism on this system, such as that from Power and Martello (1986), who criticize the system proposed by Sacks, Schegloff and Jefferson (1974) for not incorporating communicative modes outside of speech into their system. Goodwin’s (1980) research seems

² See chapter 4 for an extensive overview of Modal Density and other aspects of Multimodal Discourse Analysis

to suggest that doing so may provide evidence for a system of more constant communicative exchanges, rather than the clear taking of turns. The very last sentence of his paper actually strongly follows this path, stating ‘(...) the talk produced within a turn is not merely the result of the actions of the speaker, but rather is the emergent product of a process of interaction between speaker and hearer’ (p. 294). The current thesis will further explore this concept of interaction between both interlocutors creating interaction, rather than only one.

The current research analysed naturalistic data collected specifically for this study in an attempt to outline this process. It will specifically look at those situations where there appears to be no switch of turns, but the listener still influences the choices made by the speaker. Or perhaps more precisely: The speaker chooses to adapt his speech because of communicative actions he interprets from the listener, causing the occurrence of Inter Boundary Modulations or Cross Boundary Effects. This appears to be in contradiction with the second of Sacks, Schegloff and Jefferson points: (2) Overwhelmingly, one party talks at a time (p. 700). This is presumably largely true when one specifically considers spoken speech, as turns do seem to roughly function on a temporal level in dividing the moments of audible speech taking place. It seems far less clear however when considering communication as a whole, with all multimodal aspects of it. By accepting that speech is not the only nor the permanently superordinate mode of communication, it seems a lot less likely that communication truly only takes place by one party during their turns. Instead, a constant exchange is going on with all parties being aware that they make meaning together. Speakers search for confirmation checks and listeners provide those both during and in between turns. There is constant information flowing from one interlocutor to the other in all directions that breaks with the rules as they are proposed by Sacks, Schegloff and Jefferson. Instances where both parties are producing communicative content simultaneously are not uncommon in the form of Inter Boundary Modulations. Importantly, those instances are not limited to turn

borders, but can occur anywhere throughout a turn. The turn as a phenomenon should not be ascribed more value than being a temporal unit, during which one of the interlocutors will presumably be the main user of the mode of spoken speech, while communicative actions continues to be undertaken bidirectionally. These actions can then be both interpreted and adhered to on the spot, resulting in an online adaption of the communicative actions as taken by the speaker. Equally, the actions can be interpreted and remembered for later use, in the form of Cross Boundary Effects.

3. Empirical Methodology

In order to determine the existence of both inter-boundary modulation as well as cross-boundary effects within a naturalistic social interaction, a goal directed instructional activity was used. In pairs, participants were either responsible to provide instructions or to receive them and eventually execute a game accordingly. In half of the cases participants were allowed to use the objects associated with the activity, in the other half of the cases they were not. Fourteen Dutch university students participated in this experiment. One participant of each pair was shown a short video about a game and they were instructed beforehand that they would be asked to explain this to the other member of the pair. This process of explanation was then filmed to be used for analysis, as well as the attempt of the second participant at solving the puzzle. This resulted in a naturalistic audio-video corpus of roughly 36 minutes of instructional interaction.

3.1 Task and Protocol

The task for the current study consisted of the explanation and playing of a short game, that none of the participants had seen or played before. The pair would be brought into a quiet room where through chance, one of them was selected to act as the explainer (henceforth: E) and the other participant would then be the listener (henceforth: L). The listener would then be asked to leave the room for a few minutes, with the message that they would be called back in by the experiment leader and to not come back into the room on their own accord.

After L had left the room, the experiment leader would explain to E that they would now be shown a short video of roughly two minutes, that would explain a game to them. The full script of this video as well as an English translation can be found in Appendix A. They were told beforehand that after watching the video, the end goal of the task would be for them to explain the game to L. They were then shown the video and given the items of the game, so

they could touch and look at the objects along with the video. At this point the experiment leader would also inform them if they were part of the group that would not be allowed to use these objects during their explanation later on. They were told they were allowed to pause and/or rewind the video as much as they felt necessary. After they finished watching the video they would then be asked if they had any questions about the game. The final instruction would be that when L would eventually attempt to solve one of the puzzles, E was asked to try to interfere as little as possible, only assisting when they felt L was clearly breaking any rules or really struggled to find the correct solution. If E understood all of this, the laptop with the video would be taken away and in half of the cases, the objects of the game would also be taken away.

The experiment leader would then go outside to talk to L. They would explain to L that E was about to explain a game to them and that afterwards, L would have to try and beat this game. L would be asked to first try and let E finish the whole of the explanation and only ask questions at the end. They were allowed to talk in order to for example answer if E asked them something or if they really felt the situation called for it. L would then be led back into the room and positioned opposite E. The camera would now be switched on and E would be allowed to start their explanation, either with or without objects. L would at this point usually just listen and at the end of E's explanation ask some questions about things that were potentially unclear. After they exchanged questions and answers, L would then be given the puzzle, either from under the table or just from E's side of the table. L would now attempt to solve one of the puzzles as designated by the experiment leader. All of the puzzles used in this experiment were of the easiest variety, as indicated in the booklet that was part of the game. When L solved the puzzle, which they managed in all 7 instances, the filming would be stopped, the participants thanked for their participation and the experiment would then be over.

3.2 The game

The game used for this task is called Roadblock and is developed by the company Smart Games. The aim of the game is to block in an escaped convict between a group of buildings by correctly placing a set of police cars around the criminal. It is originally designed for just one player, with an information booklet to explain all the rules. Players will start by copying the position of both the buildings and convict from a booklet and are then tasked by having to place the police cars in the correct positions. This looks as shown in figure one:



Figure 1: An example of the game used in this experiment. On the left, a correct solution is shown. On the right is an example of one of the pages of the booklet, showing what would normally be the starting point.

As mentioned before, a player first has to copy the page of the booklet, an example of which can be seen on the right side of figure one. This is done in order to have the buildings and criminal in the correct starting positions. The player then needs to start filling in the remaining open spaces with police cars to block in the criminal and make sure they cannot escape. There are two challenges to this. The first challenge is that all police cars have to be used and eventually find a place on the board. The second and main challenge, is that the police cars have to be placed in such a way, that the convict cannot reach the sides of the board by only passing over grey squares. Consider the example in figure two. This may initially look like a correct solution, since all police cars have been used on the board.

However, as the arrows show, the convict is able to escape towards the side or the top of the board in this scenario, not being blocked by either buildings or the police cars. Only the buildings and the police cars are able to block in the criminal, which means that a solution where all the police cars fit but the red car can reach the edges of the board by only passing over the grey roads, is an incorrect solution. Important to note is that the convict is not allowed to move diagonally and therefore the picture on the left side of figure one shows a correct solution.

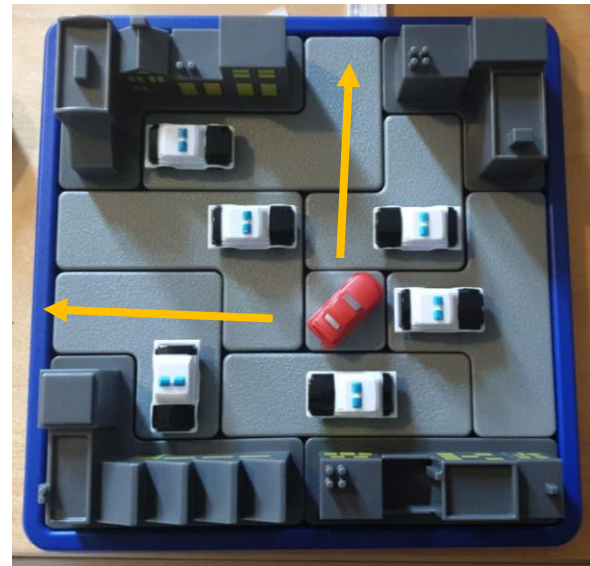


Figure 2: An example of an incorrect solution.

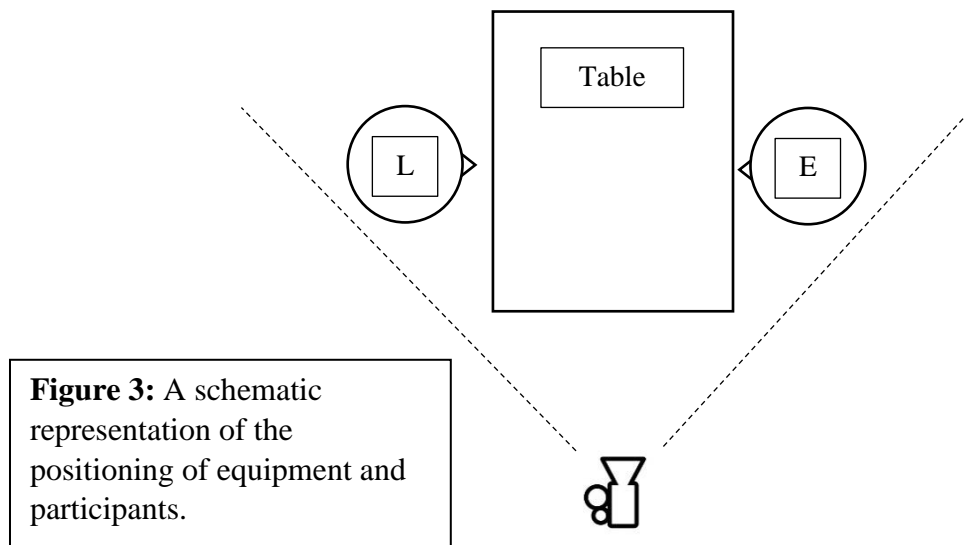
3.3 Participants

In total, fourteen Dutch university students were asked to participate in this exercise. They did not receive any form of payment or compensation for their participation. As mentioned before, the students were divided into pairs, meaning the experiment ended up with seven pairs in total. All of these students had Dutch as their native language and because of this, all interaction that was used in this experiment was done in Dutch. The participants were between the ages of 20 and 25 and all pairs consisted of two people that knew each other outside of the experimental setting.

3.4 Recording

As mentioned previously, the section that was recorded consisted of the explanation by E as well as L's attempt at solving the puzzle. On average, this resulted in a total filming time of

5.11 minutes per pair, so slightly over 5 minutes. The camera was positioned to the left of E, and to the right of L, filming the interaction effectively from the side (see figure 3).



The camera used for this recording was a Nikon D3100 which was held in a fixed position as it shown in figure three. Both participants were sitting down and filming was done from a slightly elevated position, in order to be able to properly see the table when there were objects on it or when there were gestures used at that level. The experiment leader positioned themselves behind the camera to monitor the recording and be there in case of questions or problems. They would however attempt to keep their interaction with the participants to the absolute minimum.

4. Analytical Methodology

Multimodal (Inter)action Analysis (Norris, 2004, 2011) is a methodological framework developed as a tool to help account for additional modes of communication in the qualitative analysis of real-time interaction. Multimodal (Inter)action Analysis was designed specifically with the goal of providing the methodological and theoretical tools to address the challenges faced when analysing nonverbal and verbal modes of communication together, rather than keeping both in isolation (Geenen and Pirini, 2020).

Considering all modes having their own separate materialities and structures, a singular and consistent unit of analysis was perhaps the most challenging element within such a framework. This unit of analysis had to be applicable across the whole range of different communicative modes to allow for an inductive qualitative analysis, while not implicitly allocating value to any single mode. The framework also had to be able to account and accommodate for insights that were established in studies pertaining to certain modes, such as for example gesture (McNeill, 2007) and gaze (Kendon, 1967). This unit of analysis was eventually found within Wertsch's (1991) Mediated Action Theory, which had later been adopted specifically for language by Scollon (1998): the mediated action.

The mediated action is what effectively forms the theoretical base of Multimodal (Inter)action Analysis as a framework and thereby functions as an analytical starting point. The mediated action allows for the intermingling of the individual, the sociocultural and the environmental, as they have to be recognised as being always and intricately linked and mutually influential (Scollon, 1998). Effectively, humans are therefor considered impossible to separate from the world around them, marking the 'inside' versus 'outside' distinction as problematic. Each individual is an essential element of the physical world, so they both exist as parts of each other, making a clear distinction unjustified. If you consider this point, it means that even a perceptive action is in essence bidirectionally influenced since it is the

result of an interaction with the environment. This makes even basic forms of cognition interactive, which brings up the point that essentially all actions are interactive. Every action is taken through and within the environment, making them inherently interactive.

The most significant utility of the mediated action has been highlighted in Multimodal (Inter)action Analysis via a subtle but important separation of lower-level actions and higher-level actions. This is what facilitates the application of a single unit of analysis across a large range of communicative modes. This is incredibly important, since by allowing for the use of a single unit across all modes, it ensures that an analysis cannot allocate any communicative salience to a certain mode before analysis. This is also what separates Multimodal (Inter)action Analysis from methods that take linguistic action as their analytical starting point, as this makes other modes implicitly inferior. However, rigorous analysis of communication of almost any kind will reveal such an approach as one that simply cannot hold.

4.1 The unit of analysis: Mediated actions

As mentioned previously the origins of the chosen unit of analysis for Multimodal (Inter)action Analysis lie within Wertsch's (1991) Mediated Action Theory and Scollon's (1998) subsequent Mediated Discourse Theory, both of whom were influenced by Vygotskian (1978, as qtd. in Geenen and Pirini, 2020) socio-historical psychological approaches before that. At the heart of all three are the two notions of action and mediation. Wertsch (1991) makes a critical point about the importance of action as needing analytic priority, since this ensures humans are viewed as 'coming into contact with, and creating their surroundings as well as themselves through the actions in which they engage' (p. 8). This priority and the belief that all human action is mediated makes clear that the intricate connection of 'inside' versus 'outside' is safeguarded within the basic analytical unit.

Wertsch (1991) considers the usefulness of the mediated action to lie within the fact that it maintains as much of the complexities as a part of actions as possible. Any action is invariably shaped by individual and environmental aspects of all shapes and sizes, and its nature can never be reduced to either one in isolation. It will always emerge as a result of the constant tension that exists between the two.

Scollon (1998) adopted these notions and applied them directly onto spoken language-in-interaction in his development of Mediated Discourse Theory (Scollon, 2001). His argument concerning the usefulness of the mediated action as a unit of analysis is similar to that of Wertsch, and he applied this directly to interactional sociolinguistics since he believes that discourse is best regarded as a form of social action. This idea may seem similar to the basic idea behind speech act theory, in that the utterance is classified as an action. The main use of the mediated action however, lies in the fact that it connects the social actor and mediational means in this irreducible way.

A clear example comes from the consideration of any utterance as a mediated action. By analysing the utterance as a mediated action, it is now conceptualised as being affected and influenced by all sorts of mediational means at the specific place and time it is produced. Its quality is determined by a speaker's vocal apparatus, there are lexical choices present pertaining to prosody and wording of the utterance, the utterance is perhaps part of a conversation or presentation, it is spoken during a concert or during a lecture, all spoken words have at some point been learnt within a certain socio-cultural environment. By analysing this utterance not as a linguistic unit but as a mediated action, it allows for a logical analysis in which it is permeated and affected by all sorts of related trajectories. This is what allows the mediated action to be a single useful unit of analysis, while maintaining as high a level of complexity as possible. The mediated action serves to help solve one of the most difficult aspects of human communication, which is that humans never do so through one

single mode. This unit of analysis allows Multimodal (Inter)action Analysis to analyse all modes of communication, without assigning precedence to any mode beforehand.

4.2 Lower and higher-level actions

Any form of multimodal analysis will have to face its main challenge: the variety in structure, organisation and materiality of all existing communicative modes. Spoken speech is for example auditory and its materiality is fleeting, since any form of speech only exists during its production. Structurally, spoken speech allows for the use of morphemes in all sorts of combinations, resulting in the ability to create words on the spot or combine them into sentences. The different lexical elements will, on a semantic level, each attribute their own part of meaning to the overall meaning of the utterance. When comparing this to a mode like gesture, it seems to be rather the reverse. The meaning of gesture comes from the gesture as a whole, rather than the parts it consists of (McNeill, 1992). One would not analyse the trajectory or speed of a gesture separately to determine its meaning, but rather consider the gesture as a whole. In addition to that, the basic unit of analysis within a certain mode is usually, if not always, not applicable in any other mode. There is no stroke in speech, nor are there utterances in gesture. This again is part of the value of the mediated action.

In order to further increase their usefulness, it can help to methodologically separate lower-level mediated actions from higher-level mediated actions. A lower-level action is “the smallest pragmatic meaning unit of a communicative mode” (Norris, 2004, p. 8). What this entails precisely can be radically different per mode. As mentioned before, for a gesture this would be a stroke or stroke hold. McNeill (2007, p. 33) explains the stroke as ‘the gesture phase with meaning’ and the stroke hold as “strokes in the sense of meaning and effort but occur with motionless hands”. Without a stroke or stroke hold there is no gesture, which effectively causes this to be the smallest possible pragmatic unit within the mode of gesture. These smallest pragmatic units can be conceptualised as lower-level mediated actions. They

are hard to observe in real-time interaction, as they occur simultaneously across a wide variety of modes and follow each other in very quick succession. These chains of lower-level actions can then build so-called higher-level actions, which are larger in scale and have more recognizable endings and beginnings. They are also often co-produced, meaning that lower-level actions originating from multiple social actors can add to them and build them. Higher level actions can then again build upon each other to create more higher level actions. Giving a presentation can be an example of a higher level action. Its beginning can be recognized through a word of welcome through the mode of spoken speech, accompanied by relevant gestures and posture shifts. Its ending can be similar, with a word of thanks and perhaps a wave or bow. Within such a presentation other higher level actions can exist, such as discussing a certain topic within the presentation. Through the mode of spoken language one could initiate a topic shift at the start of this part of the talk, then close off the topic at the end by shifting over to a new topic in a similar manner. In this way, lower-level actions build higher-level actions, which can then also build more higher-level actions.

It is important to keep in mind that all of these aforementioned units are heuristic units, meant only for analysis. Humans do not produce a waving gesture while uttering 'hello' through the mode of spoken speech, they simply act and greet someone familiar to them when they meet them. There are no chains of lower-level actions that create higher-level actions in real-time social interaction, only people that go about their business, act and interact during their day to day life. All of these analytical units are tools, created to help make sense of the incredibly complex processes that occur between humans over a million times each day.

One can also apply this notion of the mediated action to the everyday physical environment as it is inhabited by social actors. In their engagement with this world, social actors will allocate a certain salience to both their own engagement and perception of certain entities in the world. Entities can therefore do more than simply occupy the same space a social

actor is part of, they can also represent certain ‘frozen actions’. They can be perceived as representing some form of past mediated actions that have now been frozen in time by the presence of this object, hence the name. A painting hanging on the wall means it must have been put there by someone, therefore the presence of this painting now represents the action of it being hung on the wall. This notion therefore allows one to analyse physical entities with the same continuous focus on mediated action.

4.3 Modal configurations

All actions, lower-level and higher-level, can be analysed for the meaning they produce. The notion of modal configuration (Norris, 2009) can be used to determine the relative contribution each lower-level action adds to the meaning produced by a higher-level action. This notion thus refers to the hierarchical organisation that all relevant modes have within a particular higher-level action. In order to determine this, the first step is to assess the meaning of a certain higher-level action. Then it needs to be analysed which lower-level action is hierarchically most important to the meaning of this higher level action. This type of analysis can then reveal which modes have higher levels of salience within certain higher-level actions, creating a hierarchical structure of modal contributions for each lower-level action that is part of any higher-level action.

4.4 Modal density: the foreground-background continuum

Attention and awareness are two related concepts pertaining to the way in which people will respond to and acknowledge the presence of some actions, while ignoring or less overtly acknowledging others. This produces a focus of attention, which then imminently diminishes for actions that are less and less closely related to this action. To illustrate this concept, Multimodal (Inter)action Analysis makes use of the foreground/background continuum (Norris, 2004). This continuum explores how certain actions can be placed further towards the

foreground of one's attention, or more towards the background, while also looking at how actions can sometimes shift between levels of focus. An example can be that of two football fans watching a match on television. Their main focus is presumably on watching the match, with a conversation about the match being somewhere in the midground of their attention. When something happens or the ball goes out of play, the conversation might shift to the foreground for a moment when they exchange opinions, only to move back further towards the background when the match continues. Social actors are very capable of having multiple foci of attention and often do. Each action will then have its own place on the foreground/background continuum. To determine where on the continuum a higher-level action lies, this model looks at the level of complexity a certain action has as well as the intensity of the actions that constitute. This is collectively called modal density and this is what determines the level of attention and awareness towards a certain action. When looking back at the previous example, the complexity is determined by looking how many lower-level actions are directed towards each of the actions. The two fans will have their posture and gaze oriented towards the tv, as well as their presence in this place and perhaps even occasional comments aimed at the players and/or officials on screen demonstrate their attention towards the football match. The speech they direct at each other as well as the occasional gaze shifts towards each other demonstrate their level of attention towards their conversation. In addition, it is then important to look at how intensely each of these actions is performed. Intensity of actions can be explained as the strength with which an action is performed, and this strength is always relative to the interactional environment. If actions are performed with high levels of intensity, this indicates high level of attention. Imagine for example the different types of speech the two fans might produce while talking to the tv, versus talking to each other.

Important to notice is the fact that this system is built around so called 'interactional attention/awareness'. Interactional in this case, refers to the attention and awareness that can

be judged from actions as they are produced by people. In interaction, social actors will try to make sense of the world by judging what others are attending to and what things they are aware of. Equally, people will attempt to demonstrate their own awareness to others to help them understand interactions better. There is therefore a clear link between what people experience and what they appear to be attending to. It is however always wise to keep in mind that these two are still separate notions. People are capable of intentional and unintentional misdirection when it comes to their attention and as an analyst it is important to keep in mind that these are always judgement calls and that there is a possibility these do not match with people's experiences.

4.5 Conclusion

This chapter has laid out the core features of Multimodal (Inter)action Analysis and the tools it provides as a methodological framework for the qualitative analysis of real-time social interaction. It allows for all intricacies and complexities of human interaction to stay intact by focussing on the mediated action as a starting point. Its use of a single unit of analysis implicitly incorporates the constant tension between the individual and their environment. It means that both have their influence on every action and also creates a situation whereby no mode has inherent primacy over any other. Furthermore, it accommodates for the wide range of communicative modes that exist while still allowing them all a single unit of analysis.

In addition, Multimodal (Inter)action Analysis provides clear methodological ways to chart the hierarchical salience of the different modes when used in the construction of higher-level actions. It allows for analysis to capture which modes fill which function within a higher-level action as well as help to configure the levels of attention as they are observable. By using modal complexity and intensity as guidelines, it allows for itself to organise what modes receive higher levels of attention versus those that receive lower levels of attention. The mediated action can function as a theoretically motivated notion and a methodological

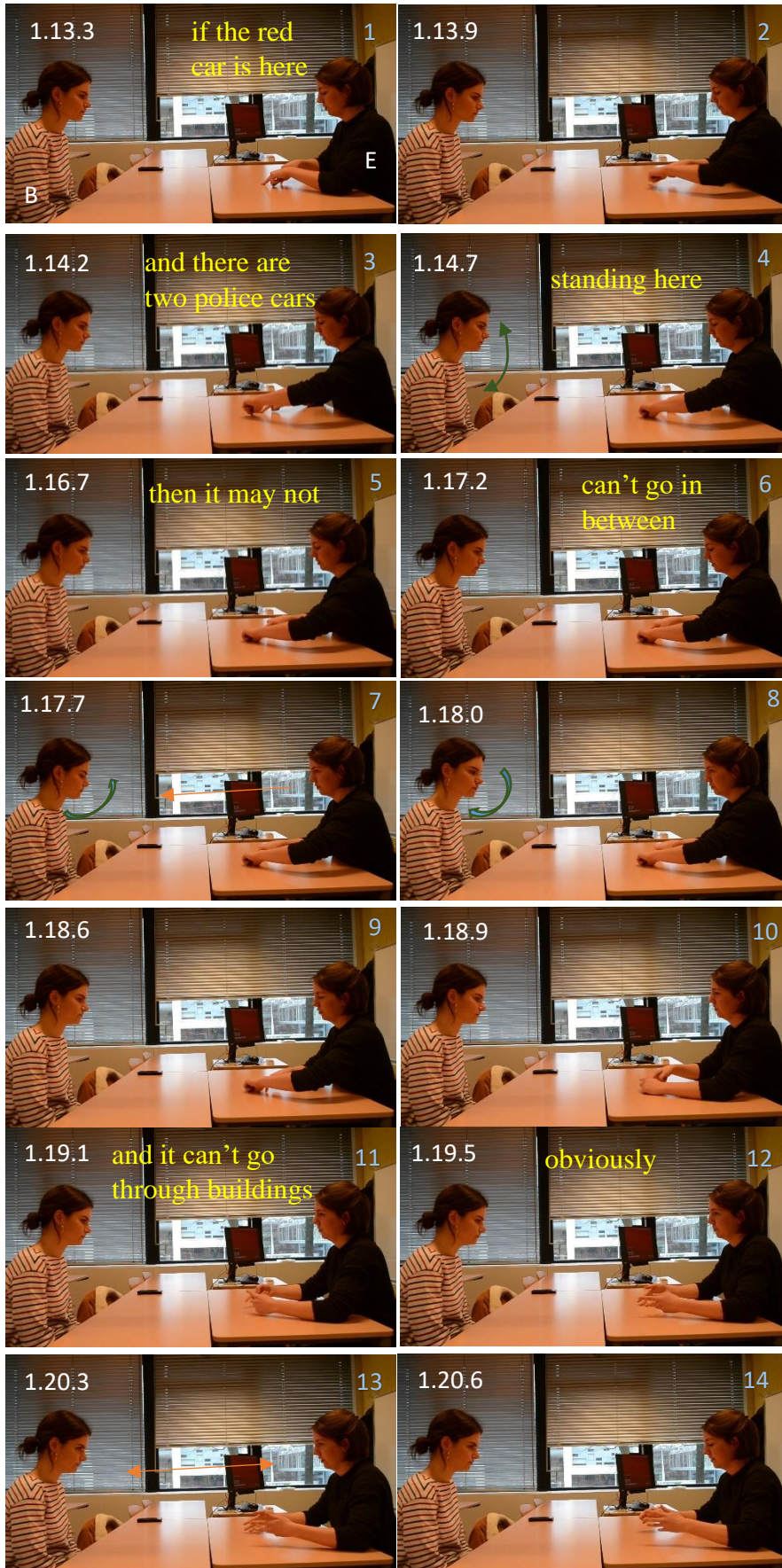
starting point, allowing for the analysis to stay fully neutral and very well suited to any empirical research concerning social interactions (Geenen and Pirini, 2020).

5. Analysis

The analysis details three representative samples from the collected data set. Each sample provides an example of either an Inter Boundary Modulation, a Cross Boundary Effect, or both. The exact material composition of each of these modulations and/or effects can be as widespread as there are modes. Their effects are potentially equally diverse, however this research presents three major and more common effects that can be had by Inter Boundary Modulations. Each of the data samples shows the occurrence of one of these effects. In the first sample, an IBM is shown that results in the shortening of a modal configuration, called an Adaptation of Production Length (APL). The second sample presents an IBM that results in a change in modal configuration, called a Shift of Mode Use (SMU). This sample also shows how an IBM can cause the occurrence of a Cross Boundary Effect. An IBM that instigates such an effect is called a Cause of Cross Boundary Effect (CCBE). The final sample will again show two examples of IBM's, one APL and one CCBE. Each of these figures will thus provide evidence towards the existence of IBM's, which are in turn instances of communicative interactions unfolding within classical turn boundaries. They are all representative samples, providing prototypical examples of something far more pervasive in the dataset as a whole.

5.1 Figure 1 – Adaptation of Production Length (APL)

This first excerpt shows an instance of Adaptation of Production Length (APL). This form of Inter Boundary Modulation (IBM) occurs when the speaker is led to believe that their communicative intention has either already been understood and their modal configuration can be scaled down or cut short, or they believe they have failed in realizing their communicative intention and feel enticed to expand their modal configuration.





In this nine second excerpt, E explains one of the more complex aspects of the game, specifically the rule that the criminal is not allowed to pass through two police cars diagonally. This is accomplished multimodally but specifically done through mainly the modes of speech, gesture and gaze. Traditional conceptualization of the turn-at-talk defines that frame 1 to 16 show only a single turn. This means the interaction as described in this section below took place right in the middle of a traditional turn-structure, during this specific turn at talk, it appears as though the interpretation of head movements which B undertakes, qualitatively affect the material unfolding of the gestures E undertakes, namely by motivating their discontinuity. Thus, the actual material qualities of the lower-level actions appear to alter in response to a non-verbal phenomenon which occurs on behalf of the 'traditional' listener. The actual rationale for these communicative actions is not found in its relation to the turns around it, but rather stem from the modulation as motivated by the unfolding actions within the boundaries of Es turn at talk. Thus, the material effect of the interpretation of B's head movements constitutes what is defined as an Inter-Boundary Modulation as it changes the material trajectory of the speakers material actions within this turn. Thus, within the boundaries of a single turn-at-talk, multimodal production trajectory alters and this is a result of E undertaking processes of production and perception simultaneously. Specifically, in frame 7, 8, 9 and 10 there is a concrete moment of IBM which is attributable to the communicative actions that E perceives during the multimodal production process.

In frame 1, E undertakes multiple lower level actions mainly through the modes of gesture, gaze and speech. She performs a deictic gesture with her left hand in which she

moves her index finger in a downwards trajectory, then touches the table in front of her. Her gaze seems to focus on the same location, which can be inferred both from her head positioning as well as the dropped upper eye lid, which suggests a focus point at around chest-level height. These actions occur simultaneously with E articulating the utterance ‘if the red car is here’³ through the mode of verbal language. This utterance is spoken at a relatively slow pace, with the prosody of the utterance ending with a rise in pitch. The combination of these three lower level actions suggest that E is denoting a specific potential location of the red car and that this potential location is a salient feature at the beginning of the instruction. She then continues her explanation by making a similar deictic gesture, with both hands this time (frame 2). She lifts up both hands for a moment, only to then once again move her index fingers towards the table in a downwards trajectory and place them down, physically touching the table (frame 3). She accompanies this movement by uttering ‘and there are two police cars standing here’.⁴ This utterance is pronounced even slower, with clear gaps between ‘two’, ‘police’ and ‘cars.’⁵ The prosody of the utterance is almost identical to that of the first one, with again an ending that rises in pitch. E’s gaze remains to appear focused on the position of her hands, as suggested again by the orientation of her head and dropped upper eye lid. This string of actions seems to function similarly to the previous ones, this time denoting potential positions of the police cars. This would also match with the previous conclusion of saliency, since the presence of the police cars in these positions explains part of the relevance of the previously introduced position of the criminal. Important to note is that she now opts to use two fingers rather than only one, suggesting there are multiple police cars and only one criminal. This would match with the overarching proposition she is communicating. Since she currently attempts to relay the fact that the criminal is not allowed to cross diagonally between

³ All these quotes have been freely translated from Dutch, the participants’ native language.

⁴ She says this fairly slowly while not moving her hands, hence the two second opening in frames.

⁵ In the original Dutch sentence, these are the three final words of the sentence

two police cars, it makes sense for there to be multiple police cars versus only a single criminal. Her hands and gestures clearly take on an important role within this explanation, as they sequentially represent all important elements of the game within this explanation.

She then continues by uttering ‘then it may not.. can’t go in between’, in again the same prosodic structure with a rise in pitch at the end. This time she does speak slightly faster and also flows into the next part of speech somewhat more easily. This verbal action coincides with mimicking the described illegal movement by making an iconic gesture in which she slides the index finger of her right hand along the table behind her left hand, which remains touching the table in an upright position as it was positioned previously (frame 5 and 6). Still her gaze seems to focus on her own hands and gestures, with her head orientation facing this way as well as the upper eye lid remaining dropped. E shows she is limited by the fact that she only has two hands and thus has to make this iconic gesture with one of the hands that previously represented one of the police cars. She does leave the other index finger planted firmly on the table, perhaps to relay that this finger has not changed in function and still denotes the potential placement of a police car. This string of actions has brought the previous actions together, now making clear that E was intending to relay a certain rule from the game, specifically a type of illegal movement for the criminal. All actions leading up to this point were meant to relay the positions of all relevant items and objects which would then function as an example in which this rule was relevant. This final string of actions shows this type of illegal movement through an iconic gesture that represents this specific event and binds together the previously performed gestures and utterances.

During this whole sequence, B continuously undertakes multiple lower level actions, mostly in the modes of gaze and head movement. She often makes small head movements, sometimes moving her head up and down in short nods and at other moments shaking it from left to right in small sideward swipes. These nods seem to match moments where E is

explaining elements like locations or items, and the shakes appear to match moments where E explains actions that are against the rules. B's gaze meanwhile, is equally focussed on E's hands as E's own look, as can be judged from her head positioning, slightly dropped upper eye lid and downwards facing irises.⁶ Throughout the first few seconds of interaction, B is constantly making these small head movements, despite there being no eye contact between the two at all during this part of the

interaction. In frame 7, E undertakes a lower-level action through the mode of gaze as she initiates a gaze shift. This is visible through her change in head positioning and raising of the upper eye lid. This is immediately after she finishes her iconic gesture, her hands now remaining still on the table in their final positions. Her gaze now appears to be focussed on B's face, as E's gaze clearly moved upwards towards a position above



Excerpt Figure 1: Frame 8 and 10 show the sudden diminishing use of gesture

chest-level. E is therefore now faced with B undertaking a strong and obvious lower level action in the mode of head movement, by shaking her head from left to right a few times. E seems to respond to this by allowing herself to start finishing the explanation. The gestures E employed beforehand now seem to be considered less important, as can be seen in frame 7, 8 and 9. E simply holds her two hands in the position she ended her iconic gesture in, not making any attempt to employ them again. The timing of this means she no longer uses her hands to gesture anymore right from the moment she notices B's head movement. It takes her only a second to actually break off the gestures, as can be seen in frame 10. She allows her

⁶ Compare for example frame 12 to 13 to see a clear difference.

left hand to lose its deictic gesture and just lie down on the table, while her right hand is pulled back slightly to make a small circling movement with an outstretched index finger. She then fluidly twists her hands around so the palms of her hands face one another. Her fingers spread out slightly and she holds them up hovering slightly above the table, towards B (frame 11 and 12). These gestures appear more metaphoric, as in speech she still explains parts of the game by uttering ‘and it can’t go through buildings, obviously.’ This utterance is produced with a more flat prosodic structure, as well as at a slightly increased pace. She thus chooses to now no longer map out locations or provide examples by building them out with her hands. This is supported by the way she employs her gaze, as she no longer aims it towards her own hands, but instead looks directly at B for this part of the explanation. This can be seen from her head positioning and now lifted upper eye lid,⁷ showing how her focus point has become much higher than it was before. B is not answering this look initially, but undertakes a lower level action within the mode of gaze in the form of a gaze shift in frame 13. She lifts her head slightly and after blinking, she now looks up a lot higher with her eyes slightly more open. This is at a point where E is holding up her hands towards B with hardly any movement. After this moment with their gazes locked, E initiates a gaze shift which results in the eye contact breaking of again (frame 14), reorienting her gaze towards the table by moving her head down and lowering the upper eye lid once more. Her hands then move towards each other and clasp together, before being directed to her chest as a single unit (frame 15 and 16). At this point, E effectively stops gesturing altogether and now simply folds her hands back to her chest to perhaps indicate the end of this chain of actions.

This figure shows the higher level action of ‘explaining the interaction between the police cars and the criminal in the game of Roadblock.’ While E starts off her explanation

⁷ Note that in frame 9, E is simply blinking. Unfortunately, the timing worked out in such a way that there was no way to insert a relevant frame where she had her eyes open between the current and previous frame.

confidently, the chosen excerpt highlights one of the moments where she seems to experience some difficulty. This is first noticed in her change in approach. While so far she uses gestures mostly abstractly and fairly random, she now actively draws attention to them. As mentioned within the analysis, they are at this point very much at the foreground of the interaction and have a high modal intensity within this interaction. She makes a reference to both the criminal and the police cars being 'here' in spoken speech, while moving her index fingers downwards towards the table in a deictic gesture. She then plants them on the table by physically touching it. Since the actual cars are obviously not around, she uses this lower level action of gesture to make a spatial reference towards a potential orientation of the criminal and the police cars towards each other within the game. She also momentarily slows down her speech, causing an almost full 2 seconds of practically no interaction except for uttering the second part of her spoken sentence.

However, almost exactly when E shifts her gaze towards the face of B (frame 10) one can see E start to change her approach towards the use of gestures, their involvement seemingly diminishing in comparison to the moments before. B does not initiate a similar gaze shift which would result in eye contact, but does continue to employ the mode of head movement by shaking her head in a sideways motion. This seems to result in E into almost instantly reducing the involvement of gestures within her modal configuration. E now starts to end her proposition early. Her gestures become more metaphoric and arguably less specific, as they change from denoting potential locations and items within the game, to being more metaphorical representations of the utterances she is producing, such as 'obviously.' This word is also an interesting choice from a semantic point of view, as apparently E feels like at this point the things she explains are obvious and thus understood by B. After moving her hands together and then directing them towards her chest she appears to finish off her use of gesture altogether, and after that she ends this part of the explanation.

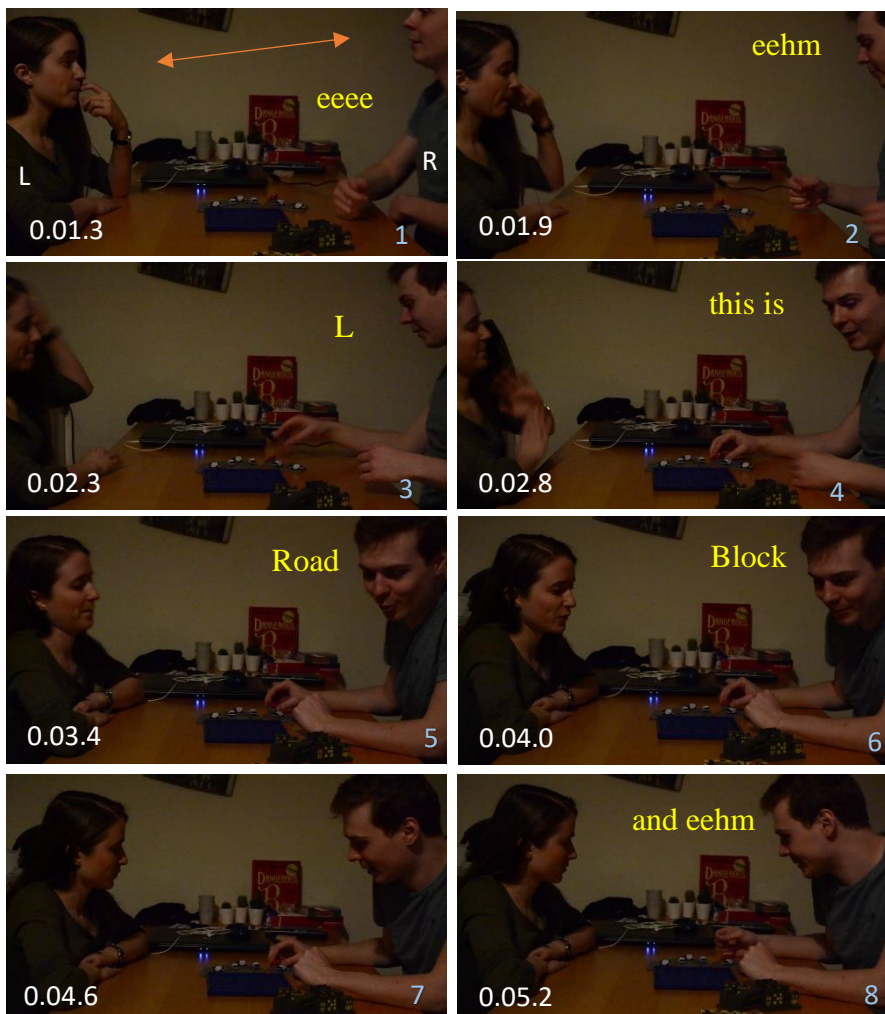
B employs the mode of head movement by making small nods and shaking her head for almost the whole of the conversation, but since E's gaze focusses down towards her own hands, she is unable to notice these actions. As soon as these lower level actions of B reach E due to E's undertaking of a gaze shift, they seem to trigger a response from E. E seems to alter the way she undertakes her actions within the mode of gesture, having them be less specific and less grand. She no longer employs them as a representation of certain relevant objects or situations outside of the conversation, but instead only uses a few small metaphoric gestures before rounding off the proposition. Her speech also changes, as the final utterance is produced with more confidence and at an increased pace. E's timing of this change in gesture usage matches up almost perfectly with her noticing B's constant head movements.

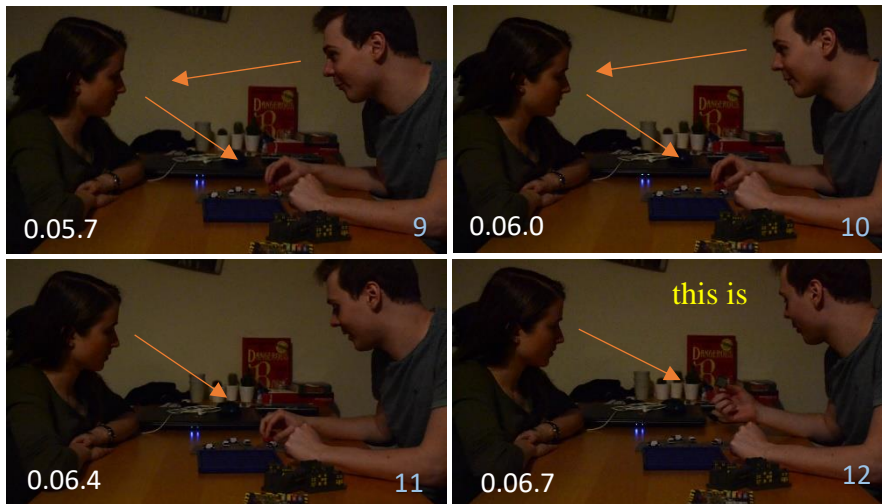
The way this scene played out appears to be a clear case of APL and with that of an IBM. The alteration in gesture use matches up almost perfectly with the moment E's gaze shift causes her to interpret B's continuous lower-level actions within the mode of head movement, even though before E notices B's undertaking of these actions her gestures were pivotal to her explanation. There also appears to be hardly any change in the co-occurring speech production, but it does still continue, adding another important rule to the game. One could argue that if E were to have continued her active gesture use to go along with her speech, nothing would have really changed within the interaction. It is also not the case that gesture is only employed when interlocutors' have their gaze directed towards the other's gesture space. Consider for example the fact that even the congenitally blind gesture when talking to other blind people (McNeill, 2007, p. 24). However, E does choose to resign her gesture use at almost the exact moment she is able to interpret B's head movements. Finally, the timing of this interaction falls right in the middle of what would traditionally constitute as a turn, making it fall in between turn boundaries. Taking this all together, there seems to be a clear case of interaction as employed by the listener directly affecting the chose modal

configurations of the speaker in real time and within the boundaries of what would classically constitute as a single turn. It does so in the form of an APL, where the length of the gesture stroke and general employment of the mode of gesture is cut short and diminished in response to head movements as undertaken by the listener.

5.2 Figure 2 – Shift in Mode Use (SMU)

This second excerpt highlights an instance of Shift in Mode Use (SMU). This form of IBM is what happens when the speaker decides to employ a different mode to use in conveying their proposition due to received input suggesting this to be the more fitting approach at this time. In this specific instance, a shift from spoken verbal language towards a foregrounding of object handling occurs as a result of perceiving a strong gaze orientation towards a specific object.





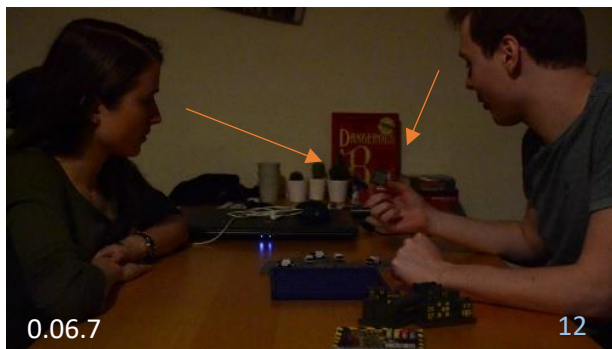
In this seven-second-excerpt, R opens his explanation. To do this, he mainly employs the modes of speech, gaze, posture and object handling. The boundaries of this turn would traditionally be identified as starting in frame 3, with R verbally articulating L's name as the first word of his utterance. The end of this articulation takes place after the current frames, namely in frame 16 as can be seen in the next section. Specifically during frame 8, 9, 10, 11 and 12, a concrete moment arises where communicative actions as undertaken by L seem to qualitatively alter the material qualities of the communicative actions as undertaken by R. It appears as if the gaze and posture orientation as employed by L, motivate the incorporation of the mode of object handling into R's modal configuration. This means that R's communicative actions appear to be more than the productive processing of a proposition or response to a previous turn. Instead, they are modulated by communicative stimuli provided by L, who would traditionally be identified as the listener. In sum, the multimodal production trajectory is altered within the boundaries of a single turn-at-talk, and this happens as a result of R's interpretations as well as his production.

The first frame shows a short moment of mutual gaze, before R utters L's name (frame 3). R allows himself a momentary stretch, but it is clear from both the participants' head positioning and direction of their irises that they are gazing towards each other's faces. In

frame 2, both participants' gaze already shifts. Their head position and dropping of the upper eye lids seems to infer that the focus of their gaze is now significantly lower, most likely towards the table that contains all the pieces of the game. The calling out takes until about the end of frame 3, 'Ehmm, L' and is spoken fairly slowly with a slight lowering of pitch towards the end. In frame 3, R can be seen to initiate an action within the mode of object handling, by moving his right hand towards the criminal on his right. Both of their postures, especially R's, then twist towards R's left, or close to the camera from the reader's perspective, but the fingers of R's right hand now close around the criminal. The buildings on R's left side of the table seem to receive the most attention from both interlocutors at this point. This seems somewhat contrastive with R's actions, as h in the meantime employs the mode of object handling and uses his right hand to pick up the criminal. Despite him now holding this, hovering slightly above the table, focus seems to be elsewhere, namely on the aforementioned buildings. R continues with naming the game, saying 'This is Roadblock', which he utters across frame 4 to 6. His speaking pace remains fairly slow and his prosodic structure shifts by employing a clear pitch rise towards the end of the utterance.

In frame 7, both participants now shift both their gazes and posture towards the other side of the table, which is the side that contains the criminal. The head movements and bodily orientation of both participants twist towards that side of the table, with their head positioning remaining low, thus now gazing towards a different element: the criminal. Interestingly, now that the criminal is presumably the actual focus of the interaction, R now actively pulls his hand back (frame 8), and refrains from employing the mode of object handling any longer. Instead, after both of the participants have reoriented their gaze and posture towards the criminal, R moves his hand with the criminal in it down towards the table, sets the criminal down and then retracts his hand back from it (frame 8). During this retraction of his hand, he also utters 'And, eehm' through the mode of speech at a more level pitch, not employing any

specific rises or falls. He does so at a fairly slow pace, especially dragging out the ‘ehhm’ and making it fairly long.



Excerpt Figure 2: Frame 10 and 12 show the contrasting approaches as taken by R

At this point, R shifts his gaze upwards towards L. His head position changes by him lifting it slightly and his upper eye lid is pulled up, suggesting his focus is now higher, towards L’s face. He inches his right hand slightly further forward again, now hovering it over the criminal while keeping his gaze focussed on L’s face, as can be seen in frame 9. Both frame 9 and 10 also show that L’s gaze remains focussed downwards, presumably towards the criminal, rather than copying R’s gaze shift to look back at

him. Her head remains tilted forward and there is no reaction in her upper eye lid, meaning she must be looking down. As R’s ‘ehhm’ lingers and his gaze is lifted towards L, R eventually appears to accept that L’s focus is going to remain on the criminal. He therefore shifts his own gaze back to the criminal as well, tilting his head down and dropping his upper eye lid again. He then re-employs the mode of object handling as he uses his right hand to pick up the criminal and lift it up towards his right, using his whole arm. He twists his lower arm upwards, meaning his hand is now below the criminal and the criminal itself sitting on top between his thumb and index finger. In the meantime he verbally continues his explanation, uttering a more confident ‘This is.’ He hereby accents the choice of focus, since the ‘This’ in his utterance clearly refers to this object he just picked up.

At the start of the explanation it becomes apparent that both gaze and posture seem to be the most important determiners of shared focus. Despite the fact that R is handling the criminal with his right hand, both his posture and direction of his gaze appear to keep the shared focus on the buildings to R's left. This is clearly followed by L, who does not seem to show any interest in the criminal just yet by also aiming both her posture and gaze towards the buildings. R's speech at this point also does not challenge this focus, since he is simply naming the game which would automatically encompass every item on the table.

When R initiates a posture shift (frame 7) by changing his bodily orientation to be directed more towards the other side of the table, the focus of both parties quickly seems to shift towards the criminal. This is also the point where R has to start his explanation proper, which he eventually does through the mode of spoken language. He lingers on this for a moment, seeking eye contact with L by aiming his gaze upwards and toward her face. L does not respond to this and keeps her own gaze aimed down towards the table. Despite actively choosing to refrain from using the mode of object handling to pick up the criminal in frame 8, after the lack of response or interaction towards him from L in frame 9 and 10, R eventually instead reemploys this mode of object handling and now lifts the criminal up completely, diverting his own gaze to it as well.

This sequence of actions as employed by R seems to illustrate a clear example of SMU. As R determines how to perform this interaction best, he uses every resource available and arguably the most important resource of all is L, the target of his whole interaction. When he starts his explanation, he first lifts his gaze from the pieces on the table and aims it towards her. He starts by using the mode of object handling by picking up the criminal at first, but currently has decided to stop doing so and initiates a gaze shift instead. He now aims his focus higher, towards L's face. This however, evokes no response at all. In fact, L's gaze stays down, looking towards the criminal still on the table. R is now faced with a choice. Since his

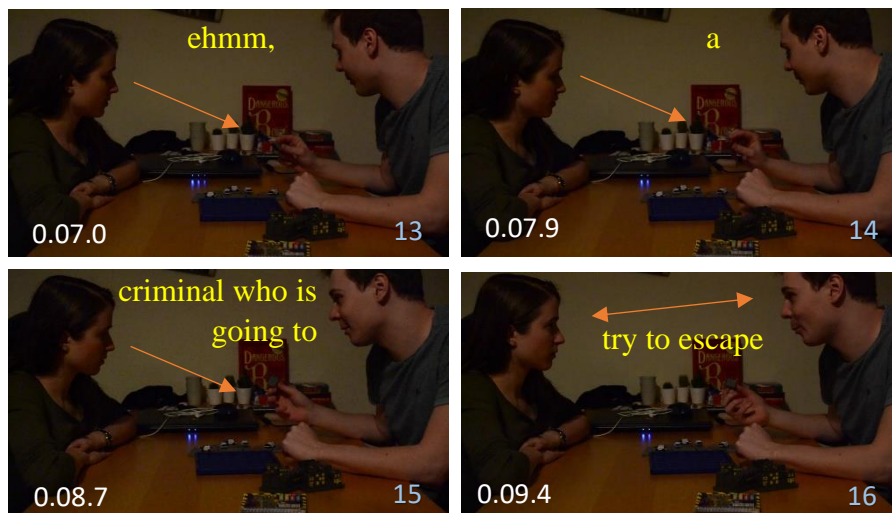
approach so far seems to not gain any response, he has to determine how to best continue. R allows the silence to hang for almost a full second, while he looks at L. He then chooses to reemploy the mode of mode of object handling and this time pick up the criminal fully, to use as the next part of his explanation. This is especially interesting since he already started this whole sequence by picking up the car, but when the focus was turned towards it he actively put it back down on the table. Something between him putting it down and the moment he picks it back up must have changed his choice of approach. It is this fact that makes it such a clear example of SMU.

Effectively, all that happens between these two moments is R initiating a gaze shift towards L that ends up not resulting in any response. Instead, R faces a listener whose gaze and posture are aimed directly and intently towards the object he just chose to put back on the table rather than keep in hand. R allows almost a second to pass before he continues his explanation, but does so by continuing the action he so consciously chose to terminate moments ago. The input he receives from L steers him back towards the object, making that the only logical way to continue his explanation. L's intent gaze on the criminal provides R with the input that shows him a clear path that ensures full attention from his audience and shifts him back towards the use of this mode. His interpretation of L's behaviour influences his decision making in his online production within the boundaries of a turn in the classical sense. It does so in the form of an SMU, where the modal configuration employed by the speaker is motivated by interpretation of communicative actions from the listener. In this specific case, the gaze orientation of the listener steered the speaker towards reemploying the mode of object handling within his modal configuration.

5.3 Figure 2 – Cause of Cross Boundary Effects (CCBE)

Figure 2 is especially interesting, since it turns out the SMU as described in the previous section, later turns out to be responsible for the occurrence of a Cross Boundary Effect (CBE).

This type of IBM is therefore called a Cause of Cross Boundary Effect (CCBE) and occurs when an action from a listener provides a speaker with information they employ later in the conversation. This could for example mean that a speaker expects a certain response at a certain point or employs a specific mode they feel the listener responded well to in the past.

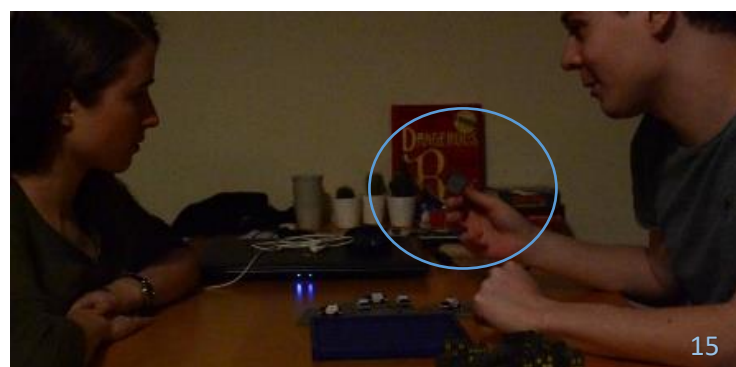


This 2 second excerpt shows the second part of a turn-at-talk as described in the previous section. During this excerpt, R is seen to be motivated in his modal configuration due to previously gained information. The material qualities of his lower-level actions cannot be directly ascribed to his communicative intentions or the material realization of the cognitive proposition which motivates his articulation. Instead, the motivations for his modal configuration stems from a moment of non-verbal communicative actions on behalf of L, that took place in the previous part of the excerpt and within classical turn boundaries. Specifically frame 15 and 16 show this moment of articulation which is attributable to communicative actions that have previously been interpreted by R.

R follows up his previous utterance of ‘This is’ with ‘ehmm, a criminal who is going to try to escape.’ He pronounces this utterance calmly, with a pitch rise at the end. At first (frame 13 and 14) he simply turns the criminal around between his fingers, seemingly at random. His thumb and index finger are positioned on the sides of the actual red car, not

touching the grey platform it is standing on. He keeps his fingers there, instead twisting his lower arm around a few times. He therefor effectively has his hand below and above the car a few times, while still holding on to it in the same manner this whole time. In frame 15, R shifts his gaze towards L in a similar manner as he did before. The upwards tilting of his head and retraction of the upper eye lid once again show the direction of his gaze here, aimed towards L's face. L's gaze initially still follows the car as R changes its orientation up and down. This is clearly visible in the footage as L's head makes small shifts and movements in synchrony with the car's movements as R tilts it up and down. When R reaches the word 'criminal' in his utterance, he lifts his gaze as previously described. At this point he also stops twisting it, instead electing to hold his hand still with the red car aimed towards his palm and his thumb and middle finger on the car's front and back, from L's perspective. He now effectively hides the actual red car from L's gaze between his fingers and keeps it there, only showing the grey bottom part of the criminal. L now no longer has anything within her field of view to look at, since R has now taken her previous focus away. As if R planned it this way, L's gaze now shifts towards R's face by tilting her head up slightly with her upper eye lid retracting, shifting her look towards R.

In the situation as described in the previous section, L's lack of response towards R's gaze shift made him choose to adapt his modal configuration to incorporate the mode of object handling. In this instance, a very similar gaze shift was employed by R, only this time there was a response to it by L. She now shifts her gaze upwards towards R, which means R now receives a response to his action, leaving no real reason to change his approach. The similarities



Excerpt Figure 2: Frame 15 shows the moment R hides the criminal in such a way to remove it from L's view

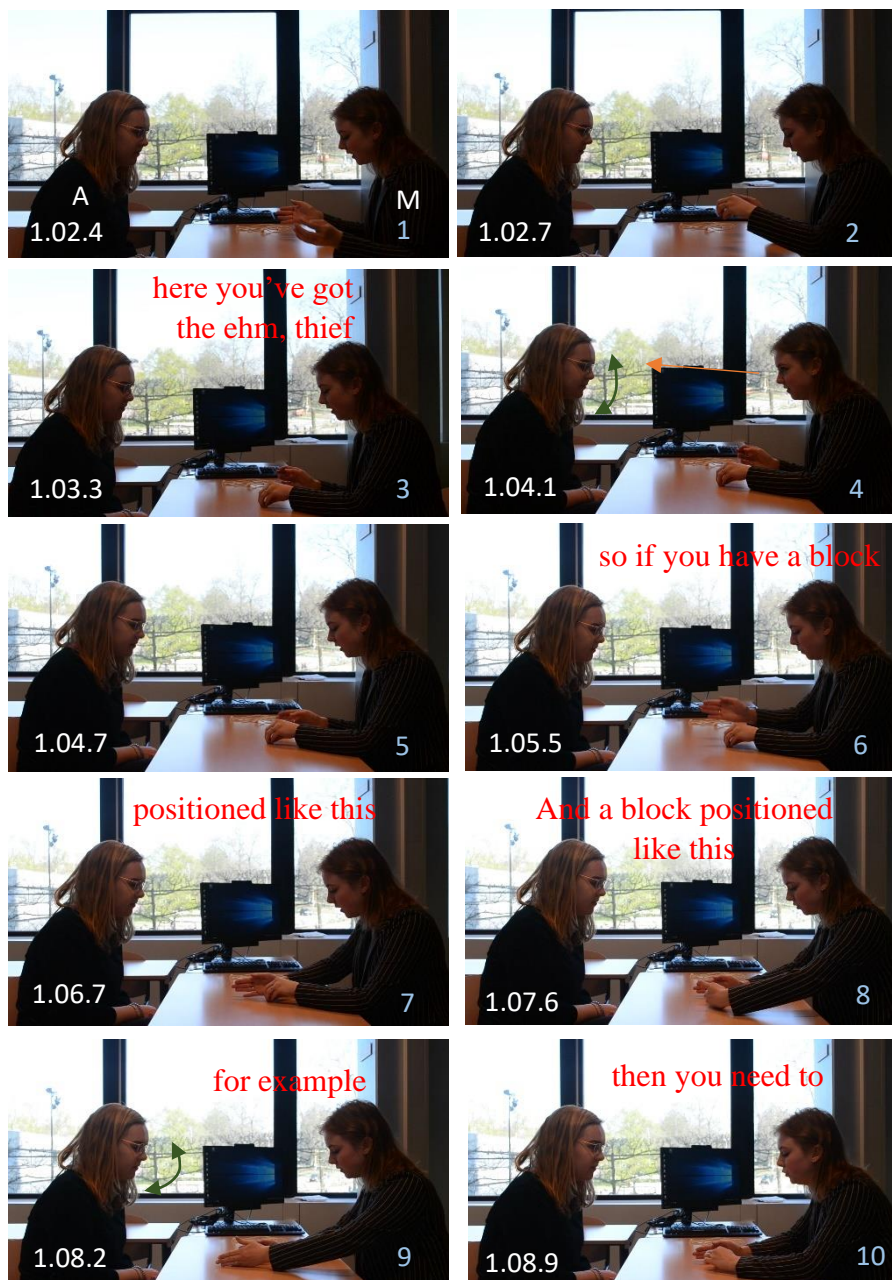
between the two situations are striking, since at first L is looking at the criminal, exactly like she was doing last time. The obvious difference is that the criminal is in R's hand this time, rather than on the table. An even bigger difference comes at the point where L initiates her gaze shift to look up towards R. At this point, the criminal is no longer just in R's hand, it is tucked away in such a manner that L can no longer see it.

Not only did the interpretations of L's behaviour influence R's online decision making at the initial moment of interpretation, they also influenced his decision making in later parts of the conversation in the form of a CBE. R finds himself in an almost identical position to the one he was in before, only this time he has much more active control over the criminal through the mode of object handling. L's gaze orientation constantly remains focussed on the criminal, which does not seem to serve R's underlying goals. When he initiates his gaze shift to direct his gaze towards L, R also undertakes a lower-level action through the mode of object handling to obscure the presence of the criminal. R seems to attempt to convey a non-verbal message indicating that the criminal is at that point not the most important part of his explanation, a proposition motivated by the fact that L's gaze orientation has previously continued to stay fixed on the criminal. These two actions combined make for L's gaze shift towards R's face, changing the outcome in comparison to the situation previously. One could wonder what would have happened if R had not obscured the criminal as such. There would have been no apparent reason for L to shift her gaze back to R, meaning R would potentially again feel the need to change approach. R's decision making has now been influenced twice by the same observation and interpretation of L's behaviour that happened within the boundaries of what classically constitutes as a single turn. The first time it was an immediate and online switch in approach in the form of an SMU, while the second time showed a more conscious use of acquired information improve future communication in the form of a CBE. The previously mentioned IBM should now be correctly defined as being both an SMU and a

CCBE, effectively changing the course of this interaction twice within a single moment of inter boundary communication.

5.4 Figure 3 – Additional evidence

This final excerpt shows additional evidence for two of the previously mentioned types of IBM. It contains both a CCBE and an APL, which are closely linked in a way similar to the situation in the previous section.





This twelve second excerpt shows the higher level action of explaining the concept of the criminal not being allowed to move between two police cars diagonally. M is the explainer in this situation, and she mainly undertakes lower level actions within the modes of gesture, gaze and spoken speech. The boundaries of the initial turn are visible in frame 3, where M initiates the explanation by starting her first utterance and seems to end in frame 14 when she finishes off this utterance. It could be argued that the closure of the turn actually only occurs in frame 19, when M receives a non-verbal confirmation from A. It seems like it could also be possible

to argue for the fact that frame 15 to 19 constitute their own new turn, but for the current proposal it is not necessarily a problem either way. At the very least, M produces communicative actions in primarily the modes of spoken speech and gesture during frame 15 to 19 that appear to be motivated by non-verbal communicative actions as they were produced by A during M's initial turn-at-talk. Her multimodal production trajectory is not simply the result of the context of previous turns, rather it is motivated by actions she interpreted during her own turn. She responds to communicative actions that were undertaken by what classically constitutes 'the listener' during her own turn-at-talk, meaning she was at that time both producing and interpreting. Her response is materialized by lengthening her modal configuration by taking another turn, or arguably even within that same turn.

In the first few frames, M undertakes a multitude of lower-level gesture actions in an attempt to visualise how the police cars and the criminal would be situated in the aforementioned scenario. Just before the start of the first frame, she determines she will more actively start to employ gesture in this part of her explanation, by asking the experiment leader: 'Can I portray this with my hands?' In the second frame she starts by raising her left hand and closes her fingers together, in preparation for a following gesture. She then moves her left hand with fingers closed together in a downwards trajectory in a deictic gesture, planting all fingers on the table with an audible thud (frame 2 and 3). During this part, her right hand remains fairly still, hovering slightly above the table. This appears similar in function to what happened in the first figure, with the gesture indicating a potential location which will prove salient in the explanation later on. Simultaneously, she produces the utterance 'here you've got, eehm, the thief', planting her fingers down as she produces the word 'here' and making a small circling gesture with her fingers touching the table while speaking the rest of the utterance. Her speech here is somewhat slow, with clear gaps. The intonation stays fairly flat, except for during the last word 'thief,' which contains a clear rise

in pitch. Her gaze is aimed downwards at first, presumably towards her own fingers. This can be judged from her head positioning, as well as the clearly visible dropped upper eye lid, which suggests a lower focus point. In frame 4, her gaze then lifts, with her head tilting back further and her upper eye lid retracting. She appears to be looking at A, who sits almost completely still opposite M, with her gaze aimed towards M's hands on the table as can be judged from her head positioning as well as her dropped upper eye lid.⁸ Despite A's gaze not shifting and remaining to be aimed downwards towards the table rather than aimed towards M, she does at exactly this moment produce a small nod by moving her head up and down in two short movements. The timing here is interesting as she does not appear to have been in a position to actually see M initiating her gaze shift, but the nods she produces do match up very well time-wise, which makes them appear very conscious.

M continues with her explanation, shifting her gaze back down by tilting her head and dropping her upper eye lid again. Apparently, nothing she interpreted triggers her towards providing an extended explanation of the situation thus far. She now starts to use her right hand. At the start of frame 5 she changes the orientation of her right hand, stretching it out more with all her fingers close to one another and pointing forward. She then shortly raises her hand before moving it in a downwards trajectory towards the table, while holding this hand upright with her pinkie finger at the bottom (frame 6 and 7). She accompanies this gesture with speech, uttering 'So if you have a block positioned like this,' which she speaks fairly fluidly and with a drop in pitch at the end. Her left hand stays in place, with all fingers pointing downwards and touching the table. The salience of the position of the fingers of her left hand now seem to return as, the relative position of these fingers to her right hand produces a situation in which her two hands now represent two relevant aspects of the game.

⁸ A's glasses make judgement of the upper eye lid slightly more difficult, however as frame 12 and 14 prove there is a clear visible difference as to whether she has them raised or dropped.

Especially as she articulates in speech that her right hand is effectively 'a block.' This means that the total sum of her actions is now starting to produce a representation of an actual in-game situation.

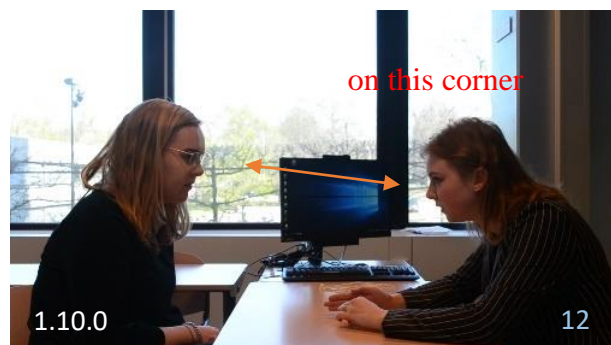
M's speech continues fairly rapidly, as she utters 'and a block positioned like this, for example.' Her intonation remains similar to before, now with two short rises in pitch towards the end, both on 'this' and 'for example'. She now allows her left hand to leave its fixed position of standing upright, orienting it in a similar position to her right hand. She holds it out flat, with her fingers close together and her pinkie finger at the bottom. In one fluid movement, she lifts up her hand, orientates it as described and then places it on the table with her pinkie finger touching the surface. She then performs a short gesture in which she drags her hand across the table towards herself while turning it at an angle, causing her fingers to momentarily point towards the palm of her right hand (frame 8) before sliding forward again to point towards A as they did before. This gesture appears to be both deictic as well as iconic, as it seems to be meant to indicate the location of the mentioned block, but it equally shows the shape of this block. M appears to be describing a block that has an angle in it, effectively an L-shaped block. Because it is impossible to set her hand in an L-shape, she chooses to make this movement to represent that shape. She therefore has now provided salience for both the shape of the block, as well as the position it takes within the described scenario and its relevant position towards the other elements she described.

A is once again very quiet during this sequence of events, but as can be seen in frame 9 she does respond to these actions as performed by M. While M is performing this gesture to represent the shape of the block, A once again undertakes a short set of head movements, moving her head up and down twice in two small nods. Both her own as well as M's gaze remain aimed fairly low, as can be judged from their head positions as well as their continuously dropped upper eye lids. In the previous situation where A produced these

movements, the timing was interesting as it coincided so well with M's gaze shift. This time however, there is no such gaze shift, yet A produces these movements all the same. This makes it somewhat unclear how conscious these movements really are, as M's gaze orientation does not seem to influence them after all.

M again continues her speech fluently and now utters 'Then you need to make sure there is a police car on this corner' (frame 10, 11 and 12). Her speech follows an intonation pattern with a rise in pitch at the end. As she speaks, she uses her left index finger to make small circling movements while touching the table, specifically in a spot she just slid over with that same left hand in her gesture to represent the L-shaped block. This is a deictic gesture, in which she is trying to relay that this block whose position she has provided before, needs to have a police car on this specific

part of itself. M undertook this small gesture action of circling on the table with her index finger, with a deictic goal of relaying the importance of having a police car on this specific part of the block. It coincides with the salience of her previous gesture, as the

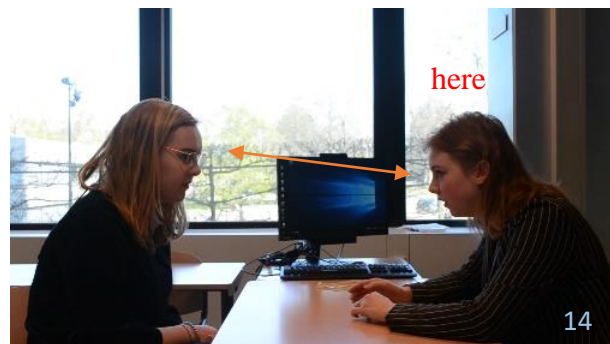


Excerpt Figure 3: Frame 12 shows the first instance both interlocutors undertake an almost identical gaze shift

position for the L-shaped block she previously established is now assumed to be understood. This current gesture in which she places her index finger down is not just meant to represent a police car standing there, but a police car standing there, on top of the previously established block. The importance of this information is further confirmed with M initiating a gaze shift in which she now lifts her head and upper eye lid to look at A's response to this information, rather than looking at her own performance of this gesture action any further. A provides a response in answering the gaze shift with a gaze shift of her own, lifting her eyelid and head to look at M's face (frame 12).

Initially, M continues with how she started. She quickly drops her gaze again, looking down towards her hands as can be seen by the momentary falling of her upper eye lid. Her accompanying utterance concludes the sentence: ‘and here.’ The drop in pitch clearly denotes the end of the sentence as well as her leaving a longer gap after finishing her verbal speech before starting a new utterance. She also provides a very similar deictic gesture to before. She slides her left hand over towards her right hand (frame 13), to then lift up said right hand while keeping its orientation of being flat and held upright. The space left behind by her right hand she then occupies with her left index finger where she makes a similar short circling gesture to indicate the importance of the position. This chain of actions is very similar to the previous one. M undertakes another deictic gesture, in which the importance of location is very much the centre of the action. She

highlights a very specific point, within the saliency of a location she established earlier, namely that of the block. She also again chooses to initiate a gaze shift as can be seen by her lifted upper eye lid and this is once again answered by A, as she does the same thing (frame 14). The pair has now shared



Excerpt Figure 3: Frame 14 shows the second instance both interlocutors undertake an almost identical gaze shift

looks on two separate occasions. On both occasions M decides to look up right after delivering vital information concerning the rules of the game. In both situations, A answers this look with a gaze shift of her own. At this point, M has verbally delivered all relevant parts of the rules. However, M does not stop her explanation here. The first time M undertook a gaze shift similar to the one she one she performs here, this did not result in any form of addition to that part of her explanation. This time however it does, as she instead chooses to continue on with her explanation. The only difference between the first instance in frame 4

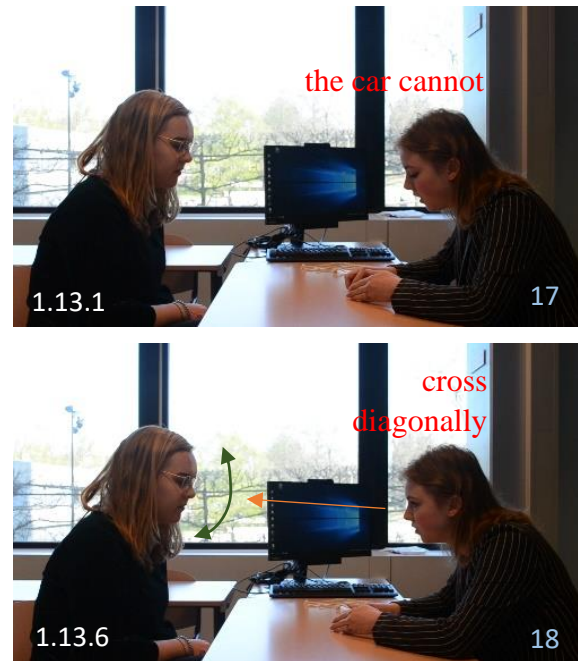
and this second instance is the way A responds. The first time she responds with small head movements, without shifting her gaze. This time, she produces these head movements but instead initiates a gaze shift. This somehow triggers M to continue her explanation, perhaps being unsatisfied by suddenly receiving a different response.

As mentioned before, M now allows a slightly longer pause between utterances. She maintains eye contact with A for a moment, but eventually shifts her gaze back down as can be seen by her upper eye lid dropping down again.

She starts a new utterance ‘So that, if its standing like this’ (frame 15). She pronounces this at the same confident pace as before, with a slight rise in pitch on the final ‘this’.

Simultaneously, she now moves both hands next to each other with outstretched index fingers touching the table. She again makes

small circling movements with her fingers (frame 16). She does this at the places on the table where she previously positioned both of the police cars separately. In the previous instance she used her left hand to make two separate deictic gestures that showed A the position of both police cars sequentially. Now she makes the decision to make a deictic gesture with both hands at the same time, making the location of the two police cars in relation to each other much more clearly visible. Crucially, she thus communicates the same proposition, but chooses a different modal configuration to relay her proposition. She continues her speech by saying ‘the car cannot diagonally cross,’ with a clear drop in pitch at the end. Simultaneously, she places her left hand vertically on the table, with her pinkie finger touching the table



Excerpt Figure 3: Frames 17 and 18 show how the outcome of the second explanation are different from the first one

(frame 17). As she utters the aforementioned speech, she also undertakes a gaze shift to again look up towards A. She slides her left hand forward alongside her right hand (frame 18), making an iconic gesture to show the described diagonal movement. As M undertakes her movement, her gaze shift allows her to now interpret the head movements undertaken by A (frame 19), after which M now does allow her explanation to end.

Effectively, M conveys the same message twice, but with a different choice in modal configuration. When looking at her choice in spoken language, it does not objectively matter whether she says 'the police cars need to be standing like this', or 'it needs to stand so, so that the car cannot cross'. During her first attempt at explaining this point, M undertook a gaze shift on two separate occasions which allowed her to interpret potential communicative actions from A. On both occasions, A's response is a very similar gaze shift. One could argue that this must be in response to M's gaze shift, considering the timing being so well matched as well as the similarity of the action. M however still decides after these two moments of eye contact to provide a second version of the explanation. This second explanation ends with M again initiating a gaze shift towards A, who does not respond with a gaze shift of her own like she did the previous two times. Instead, she undertakes the same nodding head movements she undertook earlier in the interaction, after which M now finishes this part of the explanation and continues on to the next point.

The two moments of eye contact in frames 12 and 14 result in M restating the same point in different words. The first time M shifts her gaze towards A, there was not really any response from A. A undertakes a few head movements, but there was no reason for M to assume these were meant to her. In fact, later in the interaction A can be seen to produce such head movements again without M looking up towards her. The crucial point here however, is that presumably M interprets these head movements as some form of confirmation. Perhaps the very lack of interaction by A is what prompts M to continue, as there was no reason to

believe A has any questions or finds something to be difficult to understand. Even if M does not really attribute the head movements to anything specific, a fact is that for some reason, A decides to do something different the second time M undertakes a gaze shift to look at A. M interprets these two actions produced by A, and she sees A do two different things at two otherwise very similar moments within the interaction. This is what turns the first instance of M looking towards A into a CCBE. It was at that point where she, knowingly or unknowingly, interprets actions as they are undertaken by A. She does so not only in the middle of her turn, but she does so while still simultaneously producing herself. These interpretations later influence her in her own online production. The actual moment itself, where M's gaze shift results in her interpreting A's gaze shift as something she does not expect, results in an APL. It causes her to add to her explanation by restructuring her modal configuration in a second attempt to relay her communicative intention.

This chapter shows three instances of IBM's, all of which contain moment to moment analyses of how a speaker is influenced in the online decision making of their communicative production within classical turn boundaries by an interlocutor who is not at that moment the producer of the current turn-at-talk. It is important to realise that social actors continue to produce, interpret and adapt throughout conversation, whether it is their turn or not. The following chapter will continue to highlight the importance of this conclusion.

6. Conclusion

A Multimodal (Inter)action analysis (Norris, 2004, 2011) of an arguably simplistic, naturalistic, explanatory two-party task has revealed that communicative exchanges occur continuously and at unpredictable times, in moments that have been analysed as Inter Boundary Modulations (IBM). So far, three types of IBM have been presented and are considered to be regular material compositions of the IBM in general.

Analysis of representative samples reveals that interlocutors appear to both produce and interpret simultaneously and incorporate their interpretations within their ongoing communicative production processes. In each of the three situations shown, there are visible manifestations of the adaptation of a productive communicative action, as a result of interpretation of such actions from another interlocutor. These instances are defined as Inter Boundary Modulations (IBM) and three different versions of IBM's have been analysed. Firstly, IBM's are shown to have the potential to instigate an addition or shortening of modal configurations, which is defined as an Adaptation of Production Length (APL). Secondly, IBM's are shown to have the potential to motivate a change in modal configuration, which is defined as a Switch in Mode Use (SMU). Finally, there are instances where IBM's could provide information that only later turned out to have been interpreted and affect the current speaker, a situation called a Cross Boundary Effect (CBE).

The analysis in section 5.1 reveals that the interpretation of certain confirmations can lead to an utterance being cut short, because the belief has risen that the proposition is already understood. This is a form of APL, in which the interpretation of the listener by a speaker leads to them changing the length of their utterance, by either cutting it short or making it more extensive. Specifically, figure one shows how the interpretation of the speaker of certain head movements undertaken by her interlocutor, causes her to cut her gesture usage short. This phenomenon seems to be in line with the economy principle (Vicentini, 2003). Only if

there is reason to believe it necessary, for example by input from a listener, would a speaker expand on their proposition before even allowing them a response. If there is reason to believe adding to a proposition would no longer have any effect on the interpretation of the listener, they may want to cut it short instead, following the economy principle of keeping the amount of effort to a minimum.

The analysis in section 5.2 reveals that the focus and behaviour of the listener can steer the communicative choices of the speaker in a more specific way, namely by affecting the structuration of the modal configuration at a given site of engagement. This fragment shows an SMU, with the behaviour of the listener seemingly affecting the nature of an interlocutor's modal configuration. In this specific case, the gaze orientation as employed by the listener motivates the speaker to incorporate the mode of object handling into his modal configuration. This leads to a situation similar to those described by Pirini (2016), which describes certain higher level actions to be co-produced. His research shows how the emergence of certain actions can in fact only be the result of two individuals performing higher-level actions of their own and incorporating these within the overarching action they are producing together. They thus continuously influence each other as well as the shaping of the current higher level action of conversation, meaning the action only exists as a result of the cooperation of both parties. Both social actors influence the whole of the conversation as well as each other. In section 5.3 analysis reveals this SMU also functions as a CCBE. It provides the speaker with information he was later able to call upon in a way that influences him once more in the shaping of his modal configuration. The speaker now undertook actions within the already employed mode of object handling to influence the possibilities of the listener's gaze orientation. This leads to a different outcome of a very similar situation, where previously gained information seems to play a vital role in the choices made by the speaker, and because of that has a strong influence on the eventual outcome.

Analysis in section 5.4 revealed reveals additional occurrences of both an APL and a CCBE. For the APL, the lack of a response in the form of certain head movements leads to the speaker expanding on her communicative actions. The only reason this lack of head movements had the potential to trigger her was the fact that the listener employed them earlier and the speaker interpreted them. Had they occurred, they would have matched certain behaviours that had been previously produced by the listener. This is what makes this initial use of head movements by the listener a CCBE. An interesting difference between the analysis in section 5.4 and that of sections 5.2 and 5.3, is that there is a slight difference in how the CCBE relates to the other IBM present. In figure two, the initial SMU later turned out to also provide the information which resulted in a CBE. This meant this single action was both an SMU as well as a CCBE. In figure 3, The initial CCBE became apparent by the occurrence of the APL, which itself was now both an APL as well as the manifestation of the eventual CBE. The fact that a relatively small dataset as presented in this analysis can already reveal a multiplicity of manifestations of these effects is a testament to the scale and diversity of these instances in interaction.

The analysis clearly indicates the material reality of IBM's. It reveals situations in which a listener is able to undertake productive communicative actions within a conversation that is no different than if they were at that moment the interlocutor at turn. Speakers attend and adapt to such communicative actions as they are undertaken by the listeners, effectively acknowledging the fact that these listeners are at this moment communicating. This is how this analysis is able to form the starting point of a discussion concerning the structure of the current turn-taking system and its potential flaws, as will be done in the next chapter.

7. Discussion

The current study started with an interest in the concept of explanations, specifically speech production processes during interactions in which interlocutors were either allowed to use objects when detailing the rules of a game, versus when they were not allowed to use such objects. This led to secondary questions pertaining to how communicative decisions are made, as well as how they are made specifically in the context of explaining rules to a game. The analyses as they were carried out soon provided some interesting findings, namely the fact that these decisions and production processes were not autonomously motivated by only the explainer. This eventually became the focus of the entire thesis, with the main questions focussing on the concept of turn taking and its place in real-life conversation. This research therefor tries to answer whether turn taking in its current state is a reasonable model for natural human conversation. Specifically, it looks at the implications of the existence of IBM's as they are analysed within this thesis and the theoretical flaws that have been discovered in the underlying theory for the model, to conclude that the turn taking model as it currently stands cannot hold.

7.1 Implications for the turn-taking model

The analysis and theoretical arguments within this thesis consider multiple aspects of the turn-taking system as an organisational structure of conversation. It finds that the theoretical foundations for this system leave certain gaps that are difficult to explain, while the analysis empirically reveals instances of IBM's that are hard to fit into this structure in the first place. This thesis must therefore conclude that there are serious flaws with the turn-taking system as it is currently developed. Communication is a continuous, non-linear and collaborative process, with constantly intertwining cognitive processes. On a micro-level, the analysis within this thesis has been able to show how for example a certain gaze orientation can lead to the employment of a certain mode by an interlocutor. This fragment is representative of the

data in how it shows that these 'smaller' communicative actions are attended to by speakers. As they are producing communicative actions, they are also interpreting such actions as they are produced by listeners and speakers then allow them to feed their online communicative processes. Even actions without a goal as specific as a nod or utterance of the word 'yes' can have implications for the way speakers undertake their communicative actions. A situation as shown in figure two is therefore not just the occurrence of communicative actions, but of *interactions*. Two social actors are performing communicative actions which are then interpreted in such a way that it influences and shapes the interaction as it is unfolding. There are cognitive processes intertwining and feeding of each other, in order to form this higher-level action of having a conversation or explaining a game. Within Conversation Analysis and turn taking, communicative actions are only analysed in relation to the turns that come before or after them, but not in relation to simultaneous actions undertaken by a different interlocutor. This makes for an incomplete analysis, as the examples in chapter 5 are able to show. Turns are constructed through a collective effort from both interlocutors, which already makes it difficult to speak of turns at all, as the next paragraph will also show.

The analysis has shown the existence of multiple variants of the Inter Boundary Modulation (IBM). The true importance of this whole proposal is contained within the first part of that term: 'Inter Boundary.' It would be unfair to say that conversation analysts do not believe both interlocutors have a constant influence on the flow and content of a communicative exchange. What the current proposal claims however, is that these influences are not linked to any form of set order. Turns are to be identified as temporal units, each turn representing a moment in time of the unfolding conversation. The exact content of a turn within this new definition would need to be defined very precisely. It needs to encapsulate the fact that presumably, one party has the lead in some way or another, perhaps by being the only or most avid user of the mode of spoken language. At the same time, it should leave

room for the fact that there is still production by all other social actors involved within this conversation. It is important to keep in mind that no mode is by definition superordinate to any other mode, so there is no objective difference between uttering a sentence, and nodding for example. Both of them are communicative actions with the potential to relay some form of communicative content, whether it is big or small. This is effectively what makes it so difficult to keep up any form of system that includes turns. The multimodal nature of language means that the IBM's as they are shown in the analysis provide evidence that is effectively as strong as if two people were continuously talking at the same moment. Speakers adhere to what is communicated towards them, meaning they already consider the 'listener' to also be 'speaking'. This completely problematises the concept of the turn-at-talk, as even those who are talking do not seem to see it that way.

In light of the analytical results, it seems unlikely that a unified turn could be constructed during a specific 'planning phase' prior to its articulation. At the very least, it must be acknowledged that there is a disconnect between the planning and the resultant speech wherein the latter is not reducible to the former. Consciously or unconsciously, all interlocutors within any conversation are constantly undertaking communicative actions and are observed and interpreted by the speaker, or the one whose turn-at-talk it is in the classical sense. The information that a speaker interprets from this process can then immediately be incorporated into their turn and modal configurations. This leads to a model of conversation in which planning of production could only be present in a very abstract sense, if it is present at all. Planning as a cognitive process which is completed in its entirety prior to the physiological production process does not appear congruent with the material nature of face-to-face conversation. At the minimum, the linear nature of planning and production is suspect and perhaps the two processes are more intimately intertwined during communicative activities.

This is in part how the current analysis supports the main thesis of the current research. Turn-taking in the classical sense faces multiple issues when mapped onto real world situations and these issues are both empirical, as shown in the analysis, and theoretical. As chapter 2 laid out, the current psycholinguistic analysis of the construction of turns is difficult to properly incorporate into general psycholinguistic theory. There appears to be an incredibly high cognitive load that seems difficult to match with what are generally considered to be acceptable limits. On top of that, there is a fairly strong dependence on the notion of completeness and the idea that this is a requirement for proper speech production. This once again seems difficult to match with the way speech is actually produced and observed in day to day situations, as naturalistic speech can often be littered with incomplete propositions and atypical syntax. If every utterance were planned out to completeness beforehand, this seems strange to be such a regular occurrence.

The structure of predicting, planning, production as presented by Levinson (2016) is simply too restrictive to fit day to day interaction as it can be observed. Instead such a representation should strive to capture how communication is a constant exchange of information through all available modes of communication, with the interpretation process being able to influence the online communication process to the very last moment. Even speakers consider the act of 'listening' to also be part 'speaking', or undertaking communicative action. The turn itself might remain a useful notion for conversation analysts, but it must be redefined to more accurately fit the Inter Boundary Modulations that can occur and influence interaction at a moment's notice.

7.2 Online Production Processing

As mentioned before, the turn-taking system is also limited by the cognitive processes that are said to steer it. The volume and nature of IBMs in the data set are in conflict with a turn-taking position which conceives of linear prediction, planning and production processes.

Additionally, it is also hard to fit within the psycholinguistic side of the debate, since it produces a cognitive load that seems way above the generally accepted limits. To account for the facts, there needs to be room for constant and quick adaptation. A model in which every communicative action is planned out in full before it is produced does not leave room for this kind of flexibility that the current paper argues for. Turns can be seen as temporal units for the organisation of the mode of spoken speech in a sequential order, but the definition would become much too narrow if one were to determine them to be no more than the production of a fully planned communicative action.

Planning could only be part of the communicative process insofar as that certain ideas and propositions are presumably considered before they are uttered, but the actual communicative actions as they are undertaken by social actors are malleable to the highest degree and until the very last moment. Certain propositions will not wait until a speaker has received a 'turn', but will simply be communicated in some form of action right away. When a listener feels that it makes sense to somehow communicate that they are following along with the speaker, they will for example nod or utter a small 'yes', without waiting for their turn. Equally, a speaker will look for clues as to how their communicative actions are received. They are expecting communicative actions such as these small confirmations or for example outings of confusion in the form of facial expressions or head movements. This input, defined within this thesis as Inter Boundary Modulations, are something they will use and will adapt to, in order to determine the most effective communicative actions that give their propositions the best possible chance of realizing their communicative intention.

The online production process is therefore presumably very closely linked to the online interpretation process. Rather than the two processes being conceptually distinct, they have to be considered two ongoing and co-occurring processes that do not alternate between each other, but rather feed each other and thus on each other. Perhaps the two processes are in

fact not even separate at all, though further research would be necessary to stake such a claim. This could also be the reason why people have the ability to seemingly switch so quickly and effortlessly between the two processes. There is no need to initiate an active switch from one task to another, they simply have to continue doing what they are already doing, only diverting their focus. This could explain why the time between two turns is so incredibly short, as there is effectively very little that has to happen within this timeframe.

7.3 High efficiency approach

Finally, there is also the theoretical argument against turn-taking that returns to a quote from Goodwin, as mentioned in chapter 2: ‘In conversation speakers are thus faced not simply with the task of constructing sentences but also with the task of constructing sentences for hearers’ (1980, p. 277). Speakers would under normal circumstances want to always give their proposition the best possible chance of realizing their communicative intention. In order to do this, it makes sense that any speaker would try to make use of all potential resources. This is something that already is apparent from the multimodal nature of human communication. Humans do not limit themselves to verbal communication, but rather make use of all sorts of channels of communicative actions. They will support their speech with gestures or attempt to relay their current focus by using their bodily orientation. Human beings seem to attempt to increase the efficiency of their communicative actions and try to give their propositions the best chance of realizing their communicative intention.

When considering this human desire to successfully realize communicative intention, it seems very ineffective to prepare a complete communicative action that is then set in stone with no way of amending it. The input that could be received from the targeted listener is potentially one of the most valuable sources of information when determining how to package a certain proposition. The listener is effectively the one who determines whether communicating the intended proposition has succeeded, thus their input into the shaping of a

proposition is perhaps even the most important resource available. The preparation of a complete proposition would mean that there is no room to incorporate any of the initial responses or communicative actions as they are undertaken by listeners, which does not seem to be in line with this human desire to efficiently relay a certain proposition. If modal configurations were instead considered to be more malleable and adaptable, this would allow for the incorporation of interpretations of listeners' communicative actions into the speakers' turn as it unfolds. This level of adaptability would raise the potential level of efficiency by making it possible for certain adaptations to happen immediately, rather than them being forced into the next turn. As can be seen in the analysis, it appears as though this level of adaptability is reflected in the occurrence of IBMs. This could therefore potentially prevent extra turns that could otherwise have been necessary, as certain repetitions, changes in approach or clarifications can be undertaken immediately instead of them having to wait until a later turn. This would then result in a more economic exchange, since it means the least amount of wasted effort. An interlocutor would no longer be forced to wait for a next turn to repair certain communicative errors, or would at least be able to sometimes prevent the need for the full restart of a turn by being able to instead adjust the current one.

7.4 Future Research

The second chapter highlighted some of the difficulties faced by the turn-taking system concerning the psycholinguistic side of the debate. The predicting, planning and production processes as they were described by Levinson (2016) were difficult to properly match with known human cognitive limits. It seems that there would therefore be a substantial place for neurophysiological research within the current field of research. The suggestion that interpretation and production are perhaps not two conceptually distinct processes is perhaps something that could be determined through this type of research. With methods such as EEG or MRI it is perhaps possible to show that in fact there is little difference between the

cognitive process as they unfold within a listener versus a speaker. This would add evidence to the theory that both interlocutors are actually continuously producing and interpreting, rather than switching between the processes. If this research could instead identify these two processes as conceptually distinct, it could equally be interesting to see how often and in what type of situations, if at all, they are triggered simultaneously. This is also why a potentially vital piece of interaction to investigate would be the type of situation analysed as an IBM. That would be the point where the current theory expects listeners to trigger cognitive functions associated with production and speakers would be expected to employ cognitive processes associated with interpretation. In sum, neurophysiological research has the potential to discover actual evidence of our brain being involved in these processes continuously, rather than it being a sequential change between the two.

In a similar angle, it could also be enlightening to see how communication is affected if interlocutors cannot constantly see and/or hear each other. The current thesis would perhaps suggest this leads to a more free-flowing form of interaction, as there is less to influence the approach taken by the speaker. On the flipside it could influence interlocutors to a more uncertain approach, as they are lacking vital input they would normally rely on. In this same vein, a more controlled experiment could be conducted in which one of the involved social actors is instructed to provide as little communicative feedback as possible when it is not their turn. This would even allow for a potential socio-linguistic angle, in which it could be attempted to test if there is a difference in certain social judgements towards people that provide little to no communicative interaction during phases where it is not their turn. This type of research could potentially shed some light onto how expected and normal this behaviour really is, helping to answer the question whether IBMs' existence within conversation is usually the norm or the exception.

Additionally, there are probably many more forms of IBM that occur with some regularity and mapping those out could help in future analyses to more precisely determine their function and occurrences. One such example could perhaps be a type of IBM that triggers a restart or repetition. This seems to be in line with the situation described by Goodwin in section 2.3, and could also help explain certain syntactic anomalies or fragment sentences that sometimes occur in natural speech, which are difficult to explain from a traditional turn-taking standpoint. The IBM that was analysed in the third figure as an APL could perhaps also be defined more sharply as a rephrasing or extension. All of these examples are potentially worthy of their own type of IBM, but the current research has not been able to conclusively analyse them as such.

This thesis has attempted to shed more light on the organisational structure of conversation and communication. It has presented arguments as to why thinking in terms of a linear structure of turns does not necessarily match reality. Communication should instead be looked at as a single and continuous process of interaction as cognitive processes intertwine and constantly influence the collective process. There is never one social actor taking one of many turns, rather there are many social actors all sharing different turns.

References

- Bögels, S., Casillas, M., & Levinson, S. C. (2016). *To plan or to listen? The trade-off between comprehension and production in conversation*. Poster presented at the Eighth Annual Meeting of the Society for the Neurobiology of Language (SNL 2016), London, UK.
- Geenen, J., Norris, S., & Makboon, B. (2015). Multimodal discourse analysis. *The International Encyclopedia of Language and Social Interaction*, 1-17.
- Geenen, J., & Pirini, J. (2020). Multimodal (Inter)action Analysis. In J. McKinley & H. Rose (Eds.), *The Routledge Handbook of Research Methods in Applied Linguistics* (pp. 488–499). New York: Routledge.
- Goodwin, C. (1980). Restarts, pauses, and the achievement of a state of mutual gaze at turn-beginning. *Sociological inquiry*, 50(3-4), 272-302.
- Grady, J. (1998). The “conduit metaphor” revisited: A reassessment of metaphors for communication. *Discourse and cognition: Bridging the gap*, 205-218.
- Halliday, M. A. K. (1978). *Language as social semiotic: The social interpretation of language and meaning*. Hodder Arnold.
- Kendon, A. (1967). Some functions of gaze-direction in social interaction. *Acta psychologica*, 26, 22-63.
- Kress, G. R., & Leeuwen, T. V. (2001). *Multimodal discourse: The modes and media of contemporary communication*. New York: Oxford University Press.
- Kutas, M., DeLong, K. A., & Smith, N. J. (2011). A look around at what lies ahead: Prediction and predictability in language processing. *Predictions in the brain: Using our past to generate a future, 190207*.

- Levinson, S. C. (2016). Turn-taking in human communication—origins and implications for language processing. *Trends in cognitive sciences*, 20(1), 6-14.
- Levinson, S. C., & Torreira, F. (2015). Timing in turn-taking and its implications for processing models of language. *Frontiers in psychology*, 6, 731.
- McNeill, D. (1992). *Hand and mind: What gestures reveal about thought*. University of Chicago press.
- McNeill, D. (2007). How Gestures Carry Meaning. In *Gesture and Thought* (pp. 22-59). Chicago: University of Chicago Press.
- Norris, S. (2004). *Analyzing Multimodal Interaction*. London: Routledge,
<https://doi.org/10.4324/9780203379493>
- Norris, S. (2009). Modal density and modal configurations: Multimodal actions. 2009). *The Routledge handbook of Multimodal Analysis*. Oxon: Routledge, 78-90.
- Norris, S. (2011). *Identity in (inter) action: Introducing multimodal (inter) action analysis* (Vol. 4). Walter de Gruyter.
- Norris, S. (2013). What is a mode? Smell, olfactory perception, and the notion of mode in multimodal mediated theory. *Journal Multimodal Communication*, 2(2), 155-170.
- Pirini, J. (2016). Intersubjectivity and materiality: A multimodal perspective. *Multimodal Communication*, 5(1), 1-14.
- Pirini, J. (2017). Agency and Co-production: A multimodal perspective. *Multimodal Communication*, 6(2), 109-128.
- Power, R. J., & Dal Martello, M. F. (1986). Some criticisms of Sacks, Schegloff, and Jefferson on turn taking. *Semiotica*, 58(1-2), 29-40.

- Recanati, F. (1986). On Defining Communicative Intentions. *Mind & Language*, 1(3), 213-241. doi:10.1111/j.1468-0017.1986.tb00102.x
- Torreira, F., & Levinson, S. C. (2015). The effects of processing and sequence organization on the timing of turn taking: a corpus study. *Frontiers in psychology*, 6, 509.
- Sacks, H., Schegloff, E. A., & Jefferson, G. (1974). A simplest systematics for the organization of turn taking for conversation. In *Studies in the organization of conversational interaction* (pp. 7-55). Academic Press.
- Schefflin, A. (1964). The Significance of Posture in Communication Systems. *Communication Theory*, 293-305. doi:10.4324/9781315080918-22
- Scollon, R. (1998). *Mediated discourse as social interaction: A study of news discourse*. London: Addison Wesley Longman Limited.
- Scollon, R., & Scollon, S. W. (2001). Discourse and Intercultural Communication. *The handbook of discourse analysis*, 538.
- Vicentini, A. (2003). The economy principle in language. *Notes and Observations from Early Modern English Grammars. Mots, Palabras, Words*, 3, 37-57.
- Wertsch, J. V. (1991). *Voices of the mind: A sociocultural approach to mediated action*. Cambridge, MA: Harvard University Press.

Appendix A

Original script instructional video:

- Welkom bij Roadblock. In dit spel ben jij als politie commandant verantwoordelijk voor het insluiten van een op de vlucht geslagen crimineel. Om dit te doen, ga je als volgt te werk.
- Roadblock bestaat uit een aantal onderdelen.
 - o Het speelbord
 - o De gebouwen van de stad
 - o De politieagenten
 - o De crimineel
 - o En het spelboekje.
- Voor je begint met spelen moet je het speelbord gereed maken. Dit doe je door een puzzel te kiezen uit het spelboekje en vervolgens de gebouwen en de crimineel te plaatsen zoals hierin staat aangegeven.
- Als het speelbord gereed is kun je beginnen. Om het spel te winnen moet je twee belangrijke doelstellingen halen. Als eerste is het uiteraard van belang dat de crimineel niet kan ontsnappen. Hiervoor moet je dus zorgen dat de crimineel niet tussen de politieauto's door van het bord af kan rijden. De crimineel kan ontsnappen via de grijze tegels. Hij kan niet schuin tussen politieauto's door en ook niet door gebouwen heen.
- Als tweede is het de bedoeling dat alle politieauto's hun plek vinden op het speelbord. Je moet dus uiteindelijk zorgen dat alle auto's passen, zonder dat de crimineel kan ontsnappen.
- En dat is dat, tijd om een puzzel uit te kiezen en veel succes!

Translated script instructional video:

- Welcome to Roadblock. In this game, you are the police chief and are responsible for surrounding an escaped convict. To achieve this, you take the following steps.
- Roadblock consists of a few parts:
 - o The game board
 - o The buildings of the city
 - o The police officers
 - o The convict
 - o And the rule booklet
- Before you start playing, you have to prepare the game board. You do this by choosing a puzzle from the booklet and subsequently placing the buildings and convict as the booklet shows you.
- When the game board is ready, you can get started. To win the game you have to meet two important conditions. Firstly, it is of course very important that the convict cannot escape. You therefore have to make sure he cannot reach the edges of the board by driving in between the police cars. The convict can escape via the grey tiles. He cannot drive diagonally between police cars, nor can he go through buildings.
- Secondly, all police cars need to find their place on the game board. This means that you need to make sure that all police cars fit on the board, without the convict being able to escape.
- And that's it! Time to choose a puzzle, and best of luck!