CO-ORDINATION OF SPEECH AND GESTURE IN SEQUENCE AND TIME: PHONETIC AND NON-VERBAL DETAIL IN FACE-TO-FACE INTERACTION

Rein Ove Sikveland

Submitted for the degree of PhD

University of York

Department of Language and Linguistic Science

June 2011

ABSTRACT

This thesis explores interactional processes during and between turns of talk, and how speakers and hearers accommodate each other in this process, using particular phonetic and non-verbal resources. It aims to pin down some of the interactional background work necessary to maintain coherence between turns, and seeks to explain how efficient turn-taking is made possible. These issues are addressed with detailed attention to both sequential and temporal aspects of the interactional process, combining Conversation Analysis, phonetic analyses and gestural micro-analyses.

Focussing on hearers' role in interaction, three hearer resources have been studied, in three separate studies: (i) phonetic characteristics of verbal responses (e.g. "mhm", "yeah"), (ii) head-nods, and (iii) gesture hold. The first study investigates how phonetic characteristics are used to signal whether two consecutive verbal responses are doing the same action, and shows how these characteristics are systematically used to project a shift in topic. The second study investigates head-nods used to display anticipation of further turn production. It shows how the precise co-extension of head-nods with the speaker's turn is relevant for securing an unproblematic transition to a next turn. Timing is also central issue in the third study, which studies instances where a speaker holds their gesture beyond the (verbal) completion of their turn and into a co-participant's turn. This is a resource for bringing forward an explicit issue in understanding, and the study shows how the timing of gesture hold with a co-participant's response is crucial to resolve this understanding.

This thesis contributes towards a better understanding of how the co-ordination of phonetic and non-verbal details shape talk as doing particular actions. It problematises how we should come to understand language, and, offering new insight into hearers' roles in interaction, it challenges the traditional distinction between speaker and hearer.

CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGMENTS

AUTHOR'S DECLARATION

The material collected for this thesis has previously been described in Sikveland, et al. (2010), regarding its use in the development of corpora and speech recognition tools. None of the analyses presented in this thesis was part of that collaboration. Chapter 4 is based on a paper given at the ICCA (International Conference on Conversation Analysis) 2010, and contains substantive material used in Sikveland (forthcoming, 2012), for Language and Speech. I am sole author for this work. Some of the examples in Chapter 6 appear in a poster presentation given at Labphon (Laboratory Phonology) 2010, by Sikveland and Ogden (2010). This collaboration supported the work process towards chapter 6, however all of the analyses and writing for this chapter was performed by myself. The analyses in chapter 5 has not previously been published or presented publically.

CHAPTER 1

INTRODUCTION

During interaction speakers and hearers are constantly dealing with boundaries, establishing for example who speaks when, and where an action has been achieved and a next one may begin. How speakers and hearers (henceforth: interactants) manage to make such transitions smooth and seemingly effortless is still a great mystery, and has received much attention in research on talk-in-interaction (e.g. Schegloff, 1996b; Ford & Thompson, 1996; Ford, 2004). A solution to this mystery has been conceptualised as *projection*, i.e. the ability for interactants to forewarn and foresee what comes next (e.g. Streeck, 1995; Auer, 2005). This means that for turn-taking to work, the speaker and hearer need to constantly be in tune with each other. Although this seems to make perfect sense, much work is left in exploring *how* such projection works, i.e. how efficient turn-taking is *made* possible. For example, how can one participant of a conversation make sure that the other participant is able to foresee where a turn is heading?

The studies in this thesis seek to provide some further insights regarding these issues. The starting point for these studies is careful attention to the interactional processes speakers and hearers go through during and between turns of talk. The fundamental idea is that detailed attention to the processes that happen during/between turns of talk is a key to understanding how transitions from one turn to a next turn are negotiated. In other words, speakers and hearers get to such a transition point collaboratively.

As to the *how* of these processes, the thesis will pay particular attention to phonetic and non-verbal resources, and how hearers (i.e. non-speakers) use these to affect the ongoing interactional processes. Before turning to the particular studies however, an example will illustrate the importance of investigating the ongoing interactional process when studying how a transition from one turn to a next is made. This example shows

how the circumstances of a response changes during the production of a turn, and that the interactants accommodate those changes.

Rita is Anne's lawyer and is interrogating her in preparation for a trial against an accused exhibitionist. A relevant boundary for them to negotiate is where one question is answered and the next one can be initiated. Although it is not clear-cut where Rita's question ends, the interactants still manage to provide for smooth transitions from one question to the next (note that there are no pauses between the lines). The lines, 01-05, are labelled according to their role in the sequence.

```
(1.1) SB0008, 444, "initially"

01    RIT:      did he look at you at all      Turn/sequence in progress
02    ANN:      m[hm ]                         Confirmation #1
03    RIT:       [ini]tially                   Further turn production
04    ANN:      yeah                           Confirmation #2
05    RIT:      what did he do                 Next turn/sequence
```

The conditions for Anne's response changes as Rita adds on a further element to her turn production, i.e. *initially* in line 03 fits in as an adverbial phrase of *did he look at you at all* in 01, and works as a continuation of the same question rather than a new question. Rita is apparently making sure that her question is specific enough, in terms of getting as comprehensive a background as possible about the events that led to the offence. By responding twice, first following *did he look at you at all*, then following *initially,* Anne displays sensitivity to these changing conditions. That is, as *mhm* in 02 does not anymore respond to Rita's question as a whole, Anne produces a second confirmation, *yeah*, in 04. Interestingly, Anne chooses a lexically different response to confirm for the second time, and perhaps Anne does so to show that she still confirms, but now also considers the change to Rita's question. Is this feature of relevance for achieving such an efficient transition?

The general working assumption in this thesis is that talk is produced according to the interactional process in which it takes part. This puts a simultaneous focus on both linguistic resources and interaction, and in this thesis I will argue that the interactional

process needs to be considered as an integral part of language production. I aim to demonstrate that:

- Interactants, in turns and sequences of turns shape their language production in a way that signals its relation(s) to previous and future productions. Consequently, the function (and meaning) of language production needs to be understood in such sequential terms.
- Interactants co-ordinate verbal and non-verbal productions (or signs) according to the action in progress, and that the relative timing of diverse signs is of consequence for successfully managing the interaction, e.g. in managing transitions between turns and sequences of turns.

The rationale and motivation behind these objectives are that the way in which speakers and hearer co-ordinate their behaviour is not only a basis for what is happening at the moment, but also forms a basis for what these momentary actions make relevant to happen next (Goodwin, 2000), e.g. in a next turn of talk. As a linguistic work, this thesis investigates the relation between language productions that are adjacent and/or co-present in time, reflecting "the pace of our most experience-near, moment-by-moment deployment of utterances, not historical time (...) but conversational time" (Enfield, 2009: 10). Enfield termed this type of analysis as *enchronic* analysis, as different from *diachronic* analysis. A central quality of such enchronic analysis is that it does not assign meaning to isolated elements, but to the observable interactional place a verbal or non-verbal element takes in the development of talk. Importantly, this approach puts the temporal unfolding of language production at the centre of the analysis. This raises interesting issues regarding how to define language, and linguistics, which will be addressed further in chapter 2.

Although both speaker's and hearer's actions are important in this thesis, the focus is on particular resources hearers use either to facilitate further turn production, or to prepare for a next turn of talk. The thesis also investigates ways in which current speakers may signal that they are hearers at the same time as they perform a speaker's action. Hearer is defined as a current non-speaker, i.e. the participant(s) of the interaction who is currently the recipient to current production of propositional

content. As will be shown, hearers' behaviour and co-ordination with speaker's conduct is continuously relevant for the interactional management. This is important, as hearers are commonly regarded as passive participants in linguistic research (Linell, 2009). Also, if the hearers are active participants, and if current hearers can perform speaker's actions at the same time, what does this imply regarding the common distinction between speaker and hearer (or listener)? Drawing on previous work by e.g. Goodwin (1979; 1981), Goodwin & Goodwin (1992), Clark (1996), Hayashi (2003a), Pickering and Garrod (2004), and Mondada (2007), this thesis will seek to provide an informed response to this question.

Three different types of verbal and non-verbal detail will be explored in three separate analysis chapters. These are:

- Phonetic resources in short verbal responses (e.g. "mhm", "yeah")
- Head-nods in alignment with current talk
- Gesture holds in orientation to co-participant's talk.

The analyses will employ Conversation Analysis (CA) as a core methodology. This means that the analytic process and findings are centred around the interactants' own displayed orientations, to each other's behaviour (linguistic and other), in the emerging talk: It is on the basis of the observable interactional consequences of an interactant's conduct that I will claim that they play a key role in managing the interaction. This kind of analysis seeks to avoid assumptions (i.e. pre-definitions) of what aspects of language production constitute particular functions (see e.g. Couper-Kuhlen & Selting, 1996; Local & Walker, 2005). The methods will be further introduced in chapters 2 and 3.

Below are summaries of the analytic chapters, and motivations for doing these studies.

- *The first analysis chapter* ("Phonetic resources for 'doing the same'", chapter 4) explores how hearers use phonetic characteristics to maintain and differentiate their actions across responses, and how such a distinction is consequential for who speaks next and whether it will be on the same sequence/topic. This differs from previous research on verbal responses, or 'back-channels' (e.g. Ward & Tsukahara, 2000; Benus, Gravano, & Hirschberg, 2007), as it focuses on a

sequence of responses rather than single, decontextualised ones, and it views hearers as active contributors to interaction. This study shows how phonetic detail in certain circumstances are used to differentiate interactional functions of response tokens, and that these distinctions are based on particular kinds of (non-lexical) phonetic relationships between consecutive response tokens, in relation to the co-participant's talk. As such, this study contributes to the research that focuses on the interrelationship between phonetic characteristics and sequence (e.g. Couper-Kuhlen, 1996; Curl, 2005; Ogden, 2006).

Whereas the first analysis chapter offers a sequential approach to how hearers accommodate and influence the interactional process leading towards a next turn, the second and third analytic chapters offer a more continuous and simultaneous approach to language production, focussing on the use of (i) head-nods, and (ii) manual gestures in co-ordination with speech. These chapters study the simultaneous events that constitute interactional, and shared, meaning. The beauty of non-verbal behaviours is that they may accompany and further elaborate verbal productions. The relationship between verbal and non-verbal behaviours has been formulated and explored in previous research, particularly on manual gestures (e.g. McNeill, 1992; Goldin-Meadow, 2003). However, it is largely unexplored how gestures take part in the interactional process, within and across turns of talk. In particular, the relevance of timing non-verbal resources with ongoing verbal productions in face-to-face interaction has not been given much attention in previous research.

- *The second analysis chapter* ("Anticipatory nodding", chapter 5) addresses hearers' use of head-nods during the speaker's production of a turn. The head-nods display understanding and anticipation of the progressing turn, and this study demonstrates how the co-ordination and extension of hearer's nodding is of crucial relevance for securing shared understanding and thereby takes part in defining what will happen next. Most previous studies on head-nods treat them as single responses (e.g. Maynard, 1987; Stivers, 2008), and not as finely co-ordinated parts of a speaker's turn production.

- *The third analysis chapter* ("Gesture hold and resolving shared understanding", chapter 6) studies how gestures play an important role in bringing explicit issues

to the surface of interaction, and then resolving that issue. This is based on a collection of instances where a manual gesture is held beyond the verbal completion of a turn and into co-participant's turn. As we will see, the co-ordination and timing of manual gesture with a co-participant's response displays whether or not understanding is achieved. This study gives a detailed account for how the co-ordination of speech and gesture is used as an interactional resource.

In summary, the fundamental idea that this thesis builds on, and exploits, is that meaning comes to life as talk unfolds, and that meaning is achieved by speakers and hearers accommodating their behaviour in certain ways, according to constraints in the unfolding talk and involving details in verbal and non-verbal conduct. In studying these processes, this thesis explores systematic ways in which speakers and hearers (or interactants) shape and co-ordinate their language production, and how this facilitates the achievement of shared understanding, and provide for efficient, pro-social transitions between turns.

There are in particular three (interrelated) types of motivations that guide the analysis, towards a better understanding of how turn-taking works:

- An interest in the interactional process, i.e. how speakers and hearers collaboratively work towards a point where they may rightfully proceed from one turn to another
- Highlighting the key role of different hearer productions in relation to the interactional process
- Focussing on phonetic and non-verbal detail, as linguistic resources with which speakers and hearers manage their interaction

These motivations will be supported in more detail in the next chapter.

1.1 Outline of the thesis

The thesis is structured to provide a background (chapter 2), and material and methods (chapter 3) first, followed by three analysis chapters (chapters 4, 5 and 6), and finally a general discussion (chapter 7). The background chapter addresses the general motivations for writing this thesis, with reference to related research and frameworks. A more specific background for the studies is provided in each of the analysis chapters. Chapter 3 presents the primary and secondary materials used in this thesis, and how the analytic work was performed using software tools, and the methodology of Conversation Analysis. This chapter also gives conventions for data presentation in this thesis. Again akin to chapter 2, only general aspects of the methodology are attended to here as more specific aspects are presented in the analysis chapters. Chapters 4, 5 and 6 (sometimes referred to as studies 1, 2 and 3, respectively) report the three studies in detail, each ending with a summary and discussion. Finally, chapter 7 will draw together the findings and discussion from the analysis chapters, and embedding them within the context of extant literature.

CHAPTER 2

MOTIVATION AND BACKGROUND

This chapter situates the thesis in relation to previous studies and frameworks, while also highlighting what its motivations are. The main motivations for doing this thesis were listed in the introduction, as involving (i) more attention to the hearer as an active participant in interaction, (ii) more attention to phonetic and non-verbal detail, and (iii) attending to the interactional processes behind efficient (and pro-social) turn-taking. These motivations will be further elaborated below.

The thesis operates on the study of language, and language use, but, as mentioned in the previous chapter, seeks a highly integrated approach to language, which includes gestures and other non-verbal resources, the temporal unfolding of talk, and interactional constraints regarding turn-taking and sequential context (see e.g. Goodwin & Goodwin, 1986, 1992; Clark, 1996; Goodwin, 2000; Mondada, 2007; Schegloff, 2007; Enfield, 2009; Linell, 2009; Streeck, 2009).

In this thesis language is conceptualised as a set of resources used and co-ordinated by interactants in order to make sense of one another. Such resources include a range of signs and structures: Spoken (or in a sign language: signed) words, the ordering of words in syntactic structures, non-verbal productions like manual gestures and head-nods, and lexical and non-lexical (e.g. intonation and voice quality) aspects of phonetic productions. All of these will be viewed as potential linguistic resources for constructing meaningful utterances, and as such this thesis follows a more usage-oriented definition of language, not restricted to the traditional study of phonology, morphology, syntax and semantics (see e.g. Clark, 1996; Goodwin, 2000; Linell, 2009).

The resources explored in this thesis are mainly non-verbal and (non-lexical) phonetic details. These resources have both been considered as forming paralinguistic characteristics of language, related to emotion and attitude (e.g. Laver, 1994), or as modifying the meaning of an utterance (e.g. Jaffe, 1987). The prefix 'para-' implies that,

although these details may be meaningful in some way, their relevance in shaping meaning can be analysed, and understood, as separate from linguistic content. In this thesis I attempt to show that phonetic and non-verbal details cannot so straightforwardly be studied separately from linguistic content, in the sense that e.g. lexical items, phonetic detail and gesture contextualise each other to perform certain actions. I shall therefore avoid using the term paralinguistic.

This is not to suggest that e.g. lexical items and gesture are no different from each other. As we will see later (section 2.3), there are several features that make speech and gesture different, including the distinction between conventional and non-conventional signs (see e.g. Enfield, 2009). For instance, although users of English may certainly find ways to signal gesturally that one wants another to hurry up, this gesture could be considered non-conventional in that the relation between form and meaning is not shared between English users. In the case of words like "quick" or "hurry" on the other hand, the form-meaning relation is shared.[1] However, although it makes sense to think in different terms about different language elements like speech and gesture, I will argue that it does not necessarily make sense to hold them separate when addressing human sense-making.

In the proceeding sections I wish to pursue only sense of language: Using some of Clark's (1996) terms, I aim to demonstrate that phonetic and non-verbal details may all be necessary in "understanding ordinary *linguistic* communication" (Clark, 1996, p. 392, original emphasis).

This chapter will present previous interactional and multimodal research which forms a foundation for the extended definition of language given above. First, in section 2.1, I will focus on how interaction forms an integrated part of language use. The same section will also highlight the relevance of paying attention to hearers: With reference to a model of speech production and comprehension commonly referred to in linguistic research, I will argue that there is a lot to be gained from providing a more dynamic, interactional approach to language production. Section 2.2 will direct the reader's

---

[1] Note however that a certain gesture may *acquire* a conventional meaning: This is certainly the case for so-called emblems, e.g. the OK sign and 'the finger' (Enfield, 2009).

attention to some important structural constraints in interaction, particularly with reference to previous findings and definitions related to turn organisation. I will suggest how this thesis may fill some gaps in this type of research, by attending to the interactional processes within and between turns. Finally, in section 2.3, I will argue for the importance of working towards a multimodal understanding of language production, by combining verbal, non-verbal and interactional analyses.

Only the aspects that are general to the studies presented in chapters 4-6 will be presented here, as each analysis chapter has its own introduction and background.

2.1 A framework for studying language based on interaction

In the linguistic tradition the production and understanding of speech has been conceptualised as two ends of a transmission system. This is represented in figure 2.A with the 'speech-chain model' as given by Denes and Pinson (1993). In such a model, a speaker (or indeed, the speaker's brain) produces a verbal message, which then leaves their mouth in the form of sound waves, and reaches the hearer's (or listener's) ears and the hearer then decodes the message into meaning. This model assumes that a listener's understanding is based entirely on a speaker's linguistic output, and that the only relevant output comes from the speaker's mouth, and not at all from the rest of the speaker's body. This rather simplistic model is in conflict with a range of empirical evidence showing that even single sentences cannot be isolated from the interactive process, which involves both speaker and hearer (e.g. Goodwin, 1979; 1981). This section provides a summary of such evidence, and will further provide a framework for understanding language on the basis of interaction. It will be argued that further behavioural research is needed to develop and elaborate such a framework.
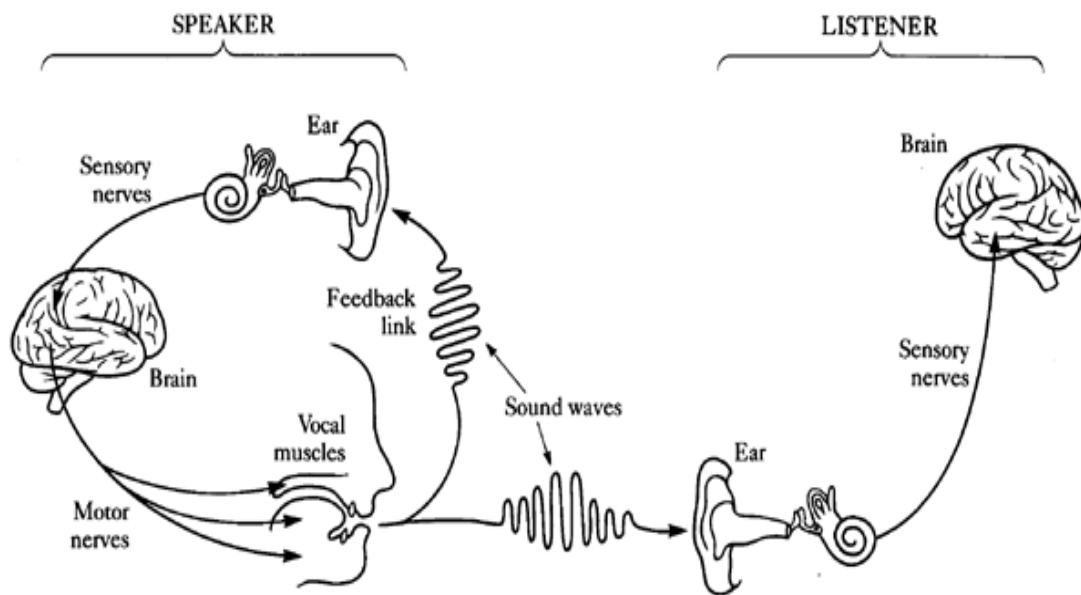
Figure 2.A The Speech Chain, by Denes & Pinson (1993)

2.1.1 Paying attention to co-present audience

There are two issues with the speech-chain model covered in this subsection, the first addressing the process of speech production, the second addressing the process of hearer understanding. In studying language production, speaker autonomy might seem like a natural starting point, as it is the speaker who produces a propositional content, and whose mouth the speech signal escapes from. And indeed this is the dominant starting point in linguistic research (Linell, 2009). But research on spoken interaction shows how hearers affect speech production during the production of a turn. One such example is that of Goodwin (1981), who demonstrated how interactants negotiate the beginning of a turn of talk by establishing mutual gaze. The basis for this finding was a collection of turn beginnings which were halted, and then restarted. Goodwin (1981) found that speakers would recurrently make these restarts when their co-participant was not gazing at them, and the speakers would proceed as the co-participant did gaze at them. This shows how details in speech production may be interactionally motivated, rather than a property of the speaker: In this case a halt in speech production is an interactionally motivated resource for securing a hearer's displayed attention/hearership.

There are also problems with viewing hearers only as a receiver in a speech chain. The speech-chain model treats hearer comprehension as linearly related to the speaker's speech production. Thus successful decoding of the message is entirely based on the speaker's signal and the hearer's ability to decode the message. Given sufficient speaking and listening conditions then, one would expect hearers to understand an utterance independently of whether or not they are addressed by the speaker. But as shown in an experimental study by Schober and Clark (1989), co-present addressees have an advantage over overhearers in terms of understanding. They argue that the key to this difference is the collaborative process, or grounding, which is a resource for co-present hearers but which is lacking for the overhearers (note that neither the co-present addressees nor the overhearers could see the speakers during the experiments).

Both of these studies show that by regarding the hearer as part of the production of a message, research gains more insight into the observable conditions that both speech production and comprehension get systematically affected by. With reference to Goodwin (1981), had hearers not been considered, an analyst might have interpreted the halts at turn beginning as entirely a speaker's problem. And contrary to what is assumed in the speech chain model, Schober and Clark (1989) shows that there is more to speech comprehension than just being within an audible range relative to the speaker. Clearly, a hearer does more than comprehending speech, and a speaker does more than producing a linguistic message.

It is the joint attention towards meaning, and action, which seems to govern language behaviour. According to Mondada (2007) talk is organised *reflexively*, in that it relies both on the production by current speaker, and on the interpretations and online analyses by hearers/recipients. This means that speakers and hearers constantly show that they are 'up to speed' on the talk in progress, displaying mutual understanding of what a language resource does in a current circumstance. In such a framework speakers and hearers are both important for creating meaning, even when only one of them does the speaking (however they do not contribute in the same way).

As an example of how speakers and hearers reflect each other during the production of talk, Goodwin and Goodwin (1992) showed how assessments are managed as an interactional achievement in conversations. Recipients to assessments get involved in the assessing activity even before the assessment is fully produced. According to these authors this is to make congruent understanding visible. Similar observations based on conversational data were made in Goodwin (1980), where a speaker appropriates some kind of appreciation from hearer, who, using vocal and non-vocal resources, displays at least some aspects of their understanding. This works to such an extent that a speaker may adjust their talk in response to recipient's actions. Again, this shows that even within single utterances speakers and hearers depend on each other to achieve meaning.

These studies show that it is not language itself that speakers orient to, but the actions speakers and hearers perform. Such findings have led several linguistic thinkers to formulate new ways of understanding, and modelling language production, putting (inter)action in the centre.

2.1.2 The centrality of action in linguistic thinking

Linell (2009) proposes a paradigmatic shift in linguistic thinking, from a monologistic and autonomous view of speech production, to a dialogistic and dynamic view. He claims that: "Linguistic items and processes are *methods* to accomplish actions, communicative projects, and to provide structure and meaning to utterances" (p. 282). So instead of viewing linguistic items (e.g. words) in themselves as meaningful, Linell (2009) suggests that linguistic work should focus on how linguistic elements are used as part of a larger project, where action is the most basic component.

In this thesis I will use the term action to describe an event in which is observable as *doing something* in the interaction. It can be a verbally produced turn of talk, which in action terms may work as a response to a question. It can be a silence gap which in action terms may disagree with a co-participant. Or it can be a pointing gesture which in action terms may direct a co-participant's attention to an object present in the room, as

part of a display of understanding. Thus, an action is some unit of behaviour which has a place following/preceding other actions, and is of potential consequence for the interaction.

In order to put this in a wider perspective, I will turn to Goodwin (2000), who states that (pp. 1489): "a theory of action must come to terms with both the details of language use and the way in which the social, cultural, material and sequential structure of the environment where action occurs figure into its organization." As for Linell (2009), language use is thought of as being at the service of action; i.e. action is based (partly) on details in language use. Goodwin (2000) also argues that we as analysts of language use need to consider a range of contextual factors, including the material surroundings in which interactants engage (see p. 1502 for an example).

Importantly, in order to gain a rich understanding of language, we cannot separate the production of signs from the contextual and interactional factors Goodwin (2000) describes. A relevant question then is what this means for thinking about language at a cognitive level. Levinson (2006) raises the issue whether there is a specific cognition for interaction which underlies all language and discourse. In this connection he proposes the 'interaction engine', conceived as the machinery that underlies human interaction, and is what makes language possible. He uses examples of how individuals from different cultures and languages may quickly find ways to communicate efficiently (for example, deaf adults with no contact with conventional sign communities quickly develop their own sign systems). The basic function of this machinery is that it 'knows' action, and actively finds ways to accommodate language accordingly. This is in a sense turning things upside down in view of a linguistic tradition. That is, instead of looking for universals in the linguistic structures (i.e. phonological, syntactic), one might find much stronger universals in the nature of human cognition and interaction. In order to build on such universals one might look for different ways of doing the same kind of action, both across and within languages.

One attempt at modelling a dialogic process from a psycholinguistic stand-point has been offered by Pickering and Garrod (2004), based on what they call 'interactive alignment'. The interactive alignment is partly automatic, where interactants align their

linguistic representations at different levels (phonological, semantic, syntactic) as the dialogue proceeds. As such the interactants have access to multiple levels of representation simultaneously. According to Pickering and Garrod (2004), this greatly simplifies language processing. This is also the major pay-off of such a model compared to more autonomous accounts. More than previous models Pickering and Garrod's (2004) model approaches an explanation of how interactants can perform common interactional phenomena, like completing a co-participant's utterance for example.

There is further potential though, in exploring what exactly constitutes and supports the interactive alignment that Pickering and Garrod (2004) propose. As mentioned above, Schober and Clark (1989) describe the additional element of common ground, as advantageous for co-present audience compared to not co-present audience (see also Clark, 1996; Clark & Krych, 2004). But what this additional element is and how it comes to be in interaction is not entirely clear. Clark and colleagues give a general description of the grounding resources as asking for confirmation, or otherwise establish the mutual belief that understanding is achieved (e.g. repair). They do not make explicit however, the exact temporal organisation, and co-ordination, with which speakers and hearers do establish such a mutual belief, and how this supports rapid turn-taking.

This thesis will maintain that a key to an understanding of these processes is by attending to details in interactional behaviour, and by conducting systematic empirical research. A danger of focussing only on a conceptual approach to action and dialogue, like in some sense Linell (2009) and Pickering and Garrod (2004) do, is that one loses the information that detailed empirical analyses might provide, and therefore fail to take into account the richness of actual interactional processes. One methodology and research area that provides such empirical accounts, while keeping (inter)action at the centre of analysis, is Conversation Analysis (CA). CA forms a strong methodological background for this thesis, and will be further introduced in chapter 3 (see also Heritage, 1989; Goodwin and Heritage, 1990; Drew, 2005; Schegloff, 2007). The next section will present some of the advances made using this method regarding the organisation of turns and turn-taking.

2.1.3 Summary

This section has presented some of the limitations in linguistic thinking, in that they view speech production as an autonomous property of a speaker, when indeed, speech production is sensitive to both the presence and the activities of the hearer. It is argued that one should instead focus on action as the central part of language production, and that further behavioural research on how speakers and hearers jointly attend to the development of action, will contribute to understanding interactional, and linguistic, processes.

2.2 Turns and turn-taking

A main motivation in this thesis is to gain a better understanding of how speakers and hearers define turn boundaries, i.e. how they establish when a unit of talk has constituted action and a next unit can begin, for example with a speaker change. This section will present an empirical basis for studying these processes, with reference to previous work on the organisation of turns and turn-taking (Sacks, Schegloff, & Jefferson, 1974; Duncan, 1974; Oreström, 1983; Lerner, 1991; Ochs, Schegloff, & Thompson, 1996; Schegloff, 2000; Ford, Fox, & Thompson, 2002), and in particular with reference to the tradition of Conversation Analysis (CA).

Based on previous research on what constitutes turn completion (2.2.1), I will argue that further work is needed in order to better understand how smooth turn-transitions are achieved, and that there is more potential attending to the collaborative processes, within and across turns, than has been exploited so far (2.2.2).

2.2.1 Turns and cues to turn completion

One might in the first instance expect that the organisation of turns and turn transitions is highly orderly, as we do them all the time, and since we rarely meet any problems in timing our talk with co-participants' contributions. Based on such observations, turns

and turn-taking has received extensive attention not only within the tradition of CA, but also in other types of interactional studies (Oreström, 1983; Campbell, 2007), and in relation to the development of dialogue systems (e.g. Edlund & Beskow, 2009).

Early attempts at formalising the turn-taking system include Duncan (1974) and Sacks et al. (1974), of which the latter has remained a classic account for the orderliness of turns in talk. I will not describe this system in detail here, but rather attend to two remaining issues with turn-taking research: (i) how to define what constitutes the recognisable units with which we organise turn-taking (e.g. Selting, 2000; Ford, 2004), and (ii) where to look for cues to turn-transition (e.g. Ford & Thompson, 1996).

Clearly, defining turns is a more complex issue than straightforwardly attributing unit categories to linguistic structures like a sentence or clause, or to intonation phrases. For example, a speaker is not necessarily finished with his/her turn with the completion of a sentence/clause, and two clauses that belong together grammatically (via a subordinate clause construction for example) may constitute separate actions in talk (Schegloff, 1996b). Instead of sticking to linguistic unit categories as such, CA researchers have attempted to develop a more action-oriented focus on what the relevant units are, and how they are organised, based on where interactants do and do not regularly initiate talk in relation to the emerging structures. Sacks, et al. (1974) introduced the concept of a turn-constructional unit (TCU), as part of the system with which conversations (and other speech-excange systems) are ordered. TCUs cover a range of unit-types which can be used to construct a turn, e.g. a lexical item, a phrase, a clause. What the different formats have in common is that they perform a recognisable action in a given context. Thus TCUs are related to but not defined by grammar; i.e. they may take different grammatical forms depending on sequential context. It is with these basic units that a speaker projects and a co-participant detects the point of completion, i.e. where a next turn can potentially start.

Sacks, et al. (1974) suggest that what defines those TCU completions where it may be relevant for a co-participant to initiate talk (i.e. transition relevance place; TRP), needs further linguistic work, an objective which has been addressed in much of turn-taking research since (e.g. Oreström, 1983; Ford & Thompson, 1996; Selting, 2000; Edlund &

Heldner, 2005; Ishi, Ishiguro, & Hagita, 2006; Barkhuysen, Krahmer, & Swerts, 2006).

Several of these studies find that when non-speakers (i.e. hearers) initiate talk, they do so at places that are characterised by the combination of certain linguistic features, mainly syntax and intonation, or prosody (Oreström, 1983; Ford & Thompson, 1996; Selting, 2000). Schegloff (1996b) noted that in cases where a construction is grammatically complete speakers may use prosody to show that one TCU does not constitute a complete turn of talk. This was confirmed by Ford and Thompson (1996) in a corpus-based analysis. They labelled a corpus of turns according to syntactic, prosodic and pragmatic completion, and found that turn transitions occurred most reliably at points where these features are combined (intonational completion was defined as final fall or final rise). Thus, syntactic completions did not alone work to provide for speaker change, but did so when co-occurring with intonational and pragmatic completions. This suggests that prosody/intonation might enhance speaker change relevance. In other words syntax seems to be nominating a possible turn completion; and the use of prosody seconds that nomination (Schegloff, 1998).

There are good reasons to be critical of the way CA research has used the term 'prosody'. It is often the case that the term prosody is referred to, without defining it any further (Local & Walker, 2004), and it is thus vague what phonetic features it is meant to include. Some phoneticians have added to a phonetically more satisfying understanding of turn-taking. For example it has been found that turn delimination correlates with loudness and tempo features, in addition to intonation/pitch features (e.g. Local, Kelly, & Wells, 1986). Further, Local and Kelly (1986) and Ogden (2001) reported on the use of glottal stop as a common resource for holding a turn, i.e. avoiding turn-transition. In other words, there are more sound-elements used in relation to turn management than those than can be described in terms of intonation.

One finding contradicts the relevance of prosody/intonation. In a study where they manipulated pitch contour and the audibility of words independently, De Ruiter, Mitterer, and Enfield (2006) found that the recognisability of lexico-syntactic structure but not pitch contour was necessary for participants to detect turn completion accurately. This result is surprising in the context of the above studies, but suggests that intonation does not have a fixed value in terms of turn completion, and is overall less

restrictive than lexico-syntactic (i.e. segmental) structure. As such this study questions the idea that transition-relevance is defined by the cluster of cues describe above. One may ask for example, whether the observation that certain prosodic features co-occur with syntactic completions necessarily means that they are the ones hearers rely on when finding an appropriate place to initiate a next turn.

Most studies on turn-taking seem to start off with the assumption that the cues to turn-taking lies in the final portions of a turn, however it is not clear whether this is actually the case. Thus the concepts of TCU and TRP may indeed be quite misleading, as many researchers focus on the definitions of their end-points rather than the process in which they become units of action (cf. Goodwin & Heritage, 1990; Schegloff, 1996b). Ford and Thompson (1996) suggest that interactants are able to detect transition relevance in advance of the actual occurrence of such a point, but that further work (e.g. linguistic, non-verbal) is needed to form a more precise specification of *how* such projection is embodied. Selting (2000) argues that a TCU is not relevant for interactants per se, but it is still an important analytic category as it is contingent on the activities that interactants are involved in. One might ask though, whether we should not be involved in analysing interactants' activities, rather than attempting to define what the units are in general. A danger with the latter, and Selting's (2000) paper, is that one might end up with an account for unit categories based on grammar after all, and lacking an account that is based on the interactional process behind the relevant units for the participants. It might be useful to shift the focus of turn-taking research from how grammatical categories define boundaries, to how interactants exploit grammatical categories to achieve turn boundaries.

2.2.2 Turn production as a collaborative processes

This thesis focuses on observable evidence that speakers and hearers do negotiate interactional boundaries, and this way providing an analysis that is more focussed on the activity at hand rather than struggling with general concepts of what a turn is. As shown by Lerner (1991) for example, hearers are able to complete a co-participant's

projected turn construction, which demonstrates that they form their actions based on anticipation and moment-by-moment orientations to meaning. Also, the work by Goodwin and Goodwin (1987, 1992) and Jefferson (1986), shows that overlapping talk are not generally a disruptive phenomenon, but a resource to co-ordinate actions with others. This shows that hearers do not just listen out for a potential completion, they take active part in the projection of a turn completion. Furthermore, this suggests that hearers are not only able to anticipate the end of a turn, but that a speaker's turn production is contingent on hearer's action. Most studies on turn-taking do not address such processes, and there is from the start a heavy focus on the speaker, and the cues this participant provides. This way there is a danger to viewing speaker's behaviour, and speaker's behaviour alone, as deterministic in terms of turn completion and transition. Sacks, et al. (1974) state that "it is misconceived to treat turns as units characterized by a division of labour in which the speaker determines the unit and its boundaries, with other parties having as their task the recognition of them" (pp. 726-727), and thus "the turn as a unit is interactively determined" (p. 727). But there is no specific mention of how hearers may affect the projection of turn completion. I will argue for the importance of hearers' actions as a rule, rather than as an exception, particularly with access to visual information. There might be more systematicity to turn-taking than has been previously been found, based on details in the hearer's conduct.

2.2.3 Summary

This section presents previous descriptions and accounts of the organisation of turns in talk. Most previous research focuses on cues to turn completion, involving syntax, intonation and phonetic detail; and not so much the interactional processes that bring forward the turn completion in the first place. This thesis seeks to elaborate the relevancies of these processes, and how interactants collaboratively use turns as a constraint, but also as a resource, to achieve their actions.

2.3 A detailed multimodal approach to talk in interaction

This thesis will perform detailed phonetic and gestural analyses within the framework of CA. As stated above, in CA there are no *a priori* assumptions about what details are important for talk in interaction (Heritage, 1989), and as such this method leaves room for discovering and empirically testing the systematic roles different types of phonetic and non-verbal details play. Also, detailed phonetic and non-verbal analysis may help understand processes in talk-in-interaction, which have not yet been described.

This section will focus on the importance of integrating non-verbal resources in the study of language and interaction (a further background on phonetic resources will be provided in chapter 4). Non-verbal resources will, in this thesis, mainly refer to manual gestures, head-nods and gaze. The first objective will be to show that non-verbal behaviour (e.g. gestures) is not additional to speech, but is an integral part of the meaning-making that speech also is a part of (2.3.1-2.3.2). Then I will present findings that show that, just like speech, non-verbal behaviour is sensitive to social-interactional contexts and processes (2.3.3). Finally, in 2.3.4, I will argue that there is further potential in studying the temporal relation between speech and non-verbal behaviour, particularly in terms of how speakers and hearers collaborate in certain interactional processes.

2.3.1 The co-ordination of speech and non-verbal behaviour

The great pay-off with investigating non-verbal behaviour and its relation to speech is to see how meaning is shaped by multiple layers of semiotic information. In this and the next subsection I will refer to studies (mainly non-CA) that show how speech and non-verbal behaviour, in particular manual gestures, are tightly linked in terms of meaning, made evident both in the production and comprehension of talk.

One type of evidence for the tight link between speech and non-verbal behaviour is their temporally co-ordinated production. In an early micro-analytic study on conversational video data, Condon and Ogston (1967) found precise correlations

between body movement and patterns in the speech stream and they called this phenomenon 'self-synchrony'. They were also intrigued to find similar patterns of synchrony between co-participating individuals, i.e. a hearer would perform bodily movements synchronised with a speaker's 'speech patterns'. Condon (1976) termed this 'interactional synchrony'.

In Condon and Ogston (1967) it was not clear what the timing relation between the speech and body movement was. However, Loehr (2007) later supported these findings, using digital equipment, and thereby economising the data collection and precise measurements. Loehr (2007) annotated intonational units and pitch accents, gestures of hands, head, and eyeblinks, and demonstrated how these recurrently form conjoined peaks, or rhythmic moments (what he referred to as 'pikes') in talk. Loehr (2007) argued that these pikes occur at regular intervals, in which we perceive rhythm. He admitted not to be able to pin down this tempo (i.e. the interval between 'pikes'), but found a common tempo in his data of about 600ms. As Condon and Ogston (1967) and Condon (1976), Loehr (2007) found this co-ordination to occur both within and across speakers. Elaborating this further, Loehr found that hearers used upcoming rhythmic moments in the speaker's talk to produce incoming talk. Based on this he proposes that humans are somehow wired to rhythmic organisation, which also corresponds to Condon's (1976) belief that there is a common neural basis for both speaking and listening.

Based on these findings it is evident that speech production is tightly and precisely timed with various forms of bodily movement, and that this might also form a basis for interactional engagement. The majority of studies on the relationship between speech and non-verbal behaviour do not elaborate their precise timing or rhythmical organisation, or their interactional relevancies. Rather, they provide a more conceptual approach to what the nature of the association between speech and body is, typically focussing on manual gestures (e.g. McNeill, 1992; Goldin-Meadow, 2003; Kendon, 2004; McNeill, 2005). Other studies provide an elaborate account of a linguistic category like prominence, typically involving facial and head movements (e.g. House, Beskow, & Granström, 2001; Beskow, Granström, & House, 2006; Swerts & Krahmer, 2008; Guaitella, et al., 2009). In my view, both these groups of studies are important because they explore and test our understanding and views of what language is. In the following

subsection I will review the research on manual gestures in greater detail, as this is more central to this thesis.

2.3.2 Manual gestures as co-expressions of meaning

According to McNeill (1992, 2005) the co-ordination between speech and gesture forms part of the evidence for a unitary bond between them, not only in production but also in underlying processes. This is in line with Kendon (1972) for example, who observes that most gestures get initiated prior to (and few if any after) their associated element in speech, and Nobe (2000) who reports that about 90 percent of representational gestures are simultaneous with co-expressive speech in this way. Even more powerful evidence for this unitary bond, is that when a speaker repairs speech that is accompanied by a gesture, he/she will initiate the same gesture again (McNeill, 1992). This has also been confirmed using Delayed Auditory Feedback (i.e. receiving auditory feedback late, which for most people disrupts speech dramatically). It has been demonstrated that the speaker in such circumstances still aligns the gesture with relevant parts of speech (McNeill, 2005). Thus this unitary bond is not broken although there is trouble in the speech production.

The tight link between speech and gesture is also evident in comprehension. For example, Habets, et al. (in press), conducted a neuroimaging (ERP; Event Related Brain Potential) study testing a listener's comprehension in different conditions where gesture-speech were simultaneous, and where gesture was delayed with certain intervals compared to speech. They found that semantic information from gesture and speech were better integrated within a certain time-frame, thus supporting the importance of co-ordination between speech and gesture. More qualitative linguistic studies of what happens when there speech and gesture lose temporal co-ordination are to my knowledge not available.

In addition to the temporal co-ordination between speech and gesture, McNeill (1992, 2005) uses *complementarity* in the gesture's relation to speech as fundamental evidence for their unitary bond. A key to this argument is demonstrating how speech

and gesture are co-expressive of meaning but also non-redundant (McNeill, 1992). Stretching this a bit further: Meaning-making would basically not be possible without gesture. This is clearly a far-fetched claim, as we *do* make sense of each other without the use and visibility of gesture: We manage well on the telephone for example. Also, one may observe that speech-accompanying gestures do not form meaning on their own as clearly as speech does. Indeed some studies suggest that listeners are less skilled at deriving meaning from gesture than from speech (e.g. Krauss, Morrel-Samuels, & Colasante, 1991). How can then gesture have as important a role as speech in language?

A common response to this (e.g. Bavelas, 1994; Goldin-Meadow, 2003; McNeill, 2005) is that the relation between speech and gesture is not a matter of assigning their relative importance, but their complementary roles in shaping meaning. Gestures are not meaningful on their own because they are always used in relation to speech. This is the way we (i.e. most of us) always see them in our everyday lives, and to a larger extent compared to speech, the meaning of gesture depends on the whole of which they are part.

A central point made by McNeill (1992, 2005) is that gestures perform aspects of meaning that speech do not, and vice versa. He introduces the distinction between gestures that match speech compared to those that are mismatched. An example of a mismatched gesture could be to describe someone walking, and use the verbal construction "walking" accompanied by a gesture that depicts a straight line. In McNeill's (1992) terms, the verbal and non-verbal elements then elaborate each other in describing that someone walked and how he/she walked; and gestures are never redundant in relation to speech. In support of this, McNeill, Cassell, and McCullough (1994) found that mismatched gestures affected listeners' comprehension, in that they create a new combination of speech and gesture that perhaps made more sense to them. This did not happen in the case of matched speech and gesture. McNeill (1992) admits that it is harder to prove the relevance of gestures that match speech compared to those that are mismatched. This is again the problem of attempting to separate speech and gesture, when they are not separate in terms of meaning (see also Alibali &

Goldin-Meadow (1993) for further evidence in favour of an integrated model of speech-gesture processing).

A further set of evidence showing that gesture is a central part of language is based on cross-linguistic studies. Duncan and McNeill (2000) studied the expression of manner in relation to verb constructions in the languages English, Chinese and Spanish, and found that these languages have systematically different ways of complementing gestures with speech, depending on where in the grammatical structure manner is coded. Thus gesture can be language-specific. But more importantly, in my view it is this sort of evidence that provides evidence for the complementarity of speech and gesture. The most relevant point of this finding is not that some gestures occur in one language and not in another, but that the way a gesture is used depends on linguistic structure. Again, these examples demonstrate how gesture forms a highly integrated part of language.

2.3.3 Speech, gesture and social processes

If there is indeed such a tight link between speech and gesture, one could expect to find that gestures, just like speech, are systematically used in the management of talk-in-interaction. With some clear exceptions, including Streeck (1995, 2009) and Mondada (2007), this is a focus that is largely lacking in gesture research, as they have mainly investigated the relations between lexical (verbal) and representational (gestures) meaning. Indeed, it seems like we have returned to the speech-chain model, where a speaker produces speech and gesture, while hearers are passive listeners. McNeill (1992, 2005) for example, does not address how this language (i.e. speech and gesture) production takes part in and gets affected by the social-interactional processes that are part of talk in dialogue (Bavelas, 2007).

There are studies that address the role social context plays on the use of gestures (e.g. Bavelas, et al., 1992; Özyurek, 2000; Furuyama, 2000). Özyurek (2000) found that speakers changed their gesture as a function of the positioning of their listeners. For example, when describing how a cat was thrown out on the street, the gestural motion for 'out' was in opposition to 'in'. This is to be expected, but interestingly, the gesture

describing 'out' changed with the relative positioning between speaker and listener. In other words the effect of audience is not only about their known presence, but also about the relation between speaker and hearer in terms of space.

This is immediately interesting in relation to the description of Enfield (2009) on Lao speakers' use of verbal demonstratives. He shows that the distinction between the verbal demonstratives *nii* and *nan* (resembling somewhat the distinction this (one) and that (one) in English, respectively), is based on fine orientations to the position of the object, the speaker and the addressee, in relation to each other. Enfield defined this distinction as 'here-space' vs. 'not here-space', and showed that what counts as 'close to' depends on the interaction. For example, if the speaker is close to the object and the addressee is not, the speaker uses a 'here-space' term (i.e. *nii*), whereas if the addressee is close to the object and the speaker is not, the speaker uses the 'not here-space' term (i.e. *nan*). In other words, both speech and gesture may be shaped by the physical surroundings and the participants' relative locations in those surroundings.

Thus, just like other parts of language (e.g. phonetic detail), gestures clearly do more than co-expressing content meaning, or lexical meaning. They are attentive to the entire social-interactional context in which meaning is shaped.

2.3.4 Summary: Bringing simultaneous-multimodal, gestural and interactional studies together

The review above gives an introduction to the study of non-verbal resources with speech, demonstrated by the temporal and co-expressive nature of speech and non-verbal resources, e.g. gesture. Thus non-verbal resources should be seen as an integrated part of language. Studies have also shown that gestures are sensitive to social and material surroundings (as are other aspects of language).

This thesis sees the opportunity to combine the simultaneous-multimodal approach and analysis of Loehr (2007) with gestural analyses, *and* with interactional analyses. The aim (particularly in studies 2 and 3) will be to study how the timing of non-verbal resources

with speech is relevant for the collaborative work that speakers and hearers do to negotiate turn boundaries. This is something few previous studies have done in a detailed and systematic way. For example, the available studies on the systematic use of gestures in a social-interactional perspective typically look at the *effect* of an interactional context, whereas what this thesis seeks to do is to establish how gesture is used to *shape* interactional context. This thesis takes advantage of the potential that lies in exploring speakers' and hearers' own real-time interactional work and negotiations, to form a better understanding of how non-verbal conduct is used as part of this process.

CHAPTER 3

MATERIAL AND METHODS

This chapter presents the material used in the thesis, and procedures for preparing and analysing the data according to the objectives of the thesis. The purpose of this chapter will be to give a general introduction to these procedures, which is relevant for all the analysis chapters. As some of the methodological descriptions are idiosyncratic to the different analysis chapters, more specific descriptions of procedures (e.g. for collection of data), will be given in the methods sections of those chapters. The details for phonetic analysis for example, will be described in chapter 4.

After giving practical and technical information about the material used (section 3.1), practical uses and implications of Conversation Analysis (CA) will be presented (section 3.2). Section 3.3 then introduces the data preparation, and how this data are presented in this thesis, i.e. transcription conventions. This focuses on representations of verbal conduct, whereas section 3.4 gives particular attention to the procedures for analysing the alignment between speech and non-verbal conduct, and conventions for representing these on paper as an addition to the verbal transcriptions. Finally, section 3.5 offers a summary of the chapter.

3.1 Material

I have used both pre-existing material, and new material collected in connection to this thesis. The main material, which I have collected myself, is a collection of conversations in Norwegian. It is predominantly this material that has been used in the studies; exclusively so in study 2 and 3. The secondary materials, used as part of the material in study 1, are pre-existing recordings of conversations in (American) English. Below is a presentation of these materials, with more detailed attention to the Norwegian

material. Procedures and technical details are given in 3.1.2-3.1.3, whereas 3.1.4 evaluates the material for the current purposes.

3.1.1 English material

One set of the English material used was 'Call Home', a collection of telephone conversations between family members and friends, collected in North America during the 1990s. The collection was made as part of a research project on speech recognition. Volunteers got to call a friend/family member (nationally or abroad) for free, for the duration of 30 minutes. The data was made available through the Linguistic Data Consortium (see http://www.ldc.upenn.edu/ for more information). Extracts from the 'Call Home' material are labelled 'CH'.

'Santa Barbara' is a collection of face-to-face conversations in American English (audio only). The conversations cover a range of activities, and include a range of social groups. See more information at http://www.linguistics.ucsb.edu/research/sbcorpus.html.

3.1.2 Norwegian material: Participants and procedure

The Norwegian material was collected at the Institute of Speech, Hearing and Science, at the Royal Institute of Technology (KTH) in Stockholm, Sweden. Examples from the Norwegian material in this thesis are identified as 'KTH-NO'. The material was collected to develop a corpus of Norwegian face-to-face interaction, as part of my role as Fellow with the 'Sound to Sense' European Research network, and to satisfy my own research interests and questions covered in this thesis. The collection was made as a subset of a larger collection for Swedish, for 'Spontal', a project that sets out to build a multimodal database for spontaneous dialogues, for studies in speech and communication (Beskow, et al., 2009). The recordings of Norwegian consisted of 30 minutes dyadic dialogues in a sound-proof recording studio, with studio-quality audio recording, and high-definition digital video recordings.

There were six participants in my Norwegian collection, one of whom participated in three separate recordings. These participants were recruited via colleagues at KTH, Stockholm. All participants were native Norwegian speakers, from south-east Norway, in and around Oslo. The details regarding the participants are presented in Table 3.A below. As is shown here, the participants were grouped in four pairs, for collection of dyadic dialogues. All the pairs were either friends or acquaintances.

Table 3.A. Overview of participants. Giving names/initials, relationship between dialogue partner, age and time spent in Stockholm/Sweden.

Dialogue pair	Participants: Assigned names (and initials)	Description of relationship	Time spent in Stockholm/Sweden	Age group
1	'Anne' (A)	Friends	Last 37 years	60+
	'Oscar' (O)		Last 18 years	60+
2	'Bengt' (B)	Friends/colleagues	Last 30 years	35-40
	'Lars' (L)		Last 8 years	30-35
3	'Sigurd' (S)	Acquaintances	On short visit	30-35
	'Lars' (L)		Last 8 years	30-35
4	'Tor' (T)	Acquaintances	Last 18 years	35-40
	'Lars' (L)		Last 8 years	30-35

The participants had stayed/lived in Sweden for varying periods of time, and some of the participants reported having developed a hybrid version of Norwegian and Swedish in their daily life. This is to be expected as Norwegian and Swedish are mutually comprehensible. Some such 'swedification' is observable in the recordings, particularly in terms of lexis, and for some speakers, also in terms of prosody. As linguistic material then, KTH-NO is perhaps best representative of 'East Norwegian spoken by speakers living in Stockholm', rather than of East Norwegian as such.

The participants were informed about the purposes of the recordings, and were willing to have the recordings be used for research purposes. They all signed a consent form, a copy (and translation) of which is in appendix A.

Each dialogue was recorded for 30 minutes, during which 20 minutes was designed as open/free conversation, whereas the last 10 minutes revolved around a given interactional task. This task was to explore the content of a wooden box present in the recording studio, and discuss the identity of the box itself, and some items that were present in the box. The box itself was a 'sugar box' ("sockerlåda" in Swedish), an instrument for cutting sugar cones into small pieces, used about a century ago. The box contained three to four objects, including models from artwork, an old-fashioned pencil sharpener, and some engineering tools. For 'Lars', who participated in three recordings, it was made sure that at least some of the items were different between the recordings.

As the participants were guided into the recording studio, they were assigned seats on either side of a small table, while head-mounted microphones, studio microphones and cameras were adjusted. This configuration is illustrated in Figure 3.A below, a still-shot of 'Anne' and 'Oscar', in preparation for their recording session. During the technical configurations the participants were presented with the aims and the structure of the recordings. Regarding the aims, the participants were told that we collected dialogue material for research on spoken language. Regarding the structure, the participants were asked to talk freely for the first 20 minutes of the recording, and then have a look into the box which was situated on the floor beside the table, and discuss the identity of the box itself and the items it contained. This task was framed as relatively free, and the participants were told that there were no requirements to keep on exploring the box throughout the entire period. In other words, they could get back to their less task-oriented conversation as it occurred natural to them. The participants were notified every 10 minutes of how much time had passed, and this way they could keep track of time. They were informed in advance that this would be done.

Figure 3.A. Configuration of recording session. 'Anne' sitting on the far side of the table, 'Oscar' sitting opposite her. To the left of the frame the two goose-neck studio microphones are visible. One of the head-mounted microphones is visible on 'Oscar's chair (not yet put on his head). Experimenter in background is adjusting one of two cameras.

3.1.3 Technical specifications

The specifications for recording equipment are presented in Table 3.B below. The audio was recorded on two sets of microphones: One set of goose-neck studio microphones, and one set of head-mounted microphones. The rationale for using double sets of microphones was that although the goose-neck microphones produce the highest-quality signal, there is much leakage in each microphone from the other subject. The inter-subject leakage is much lower in the head-set microphones.

There were two video cameras, one behind each participant, each capturing the back/side of one participant and the front of the other (see Figure 3.A above). The cameras were placed with a view of each subject that included body from their knees and up including head, at a height that was approximately level with the heads of the participating subjects, and a distance of about 1.5 meters behind the subjects to minimize interference.

Table 3.B. Summary of technical details for audiovisual recordings.

	Microphones	Audio recording	Video recording
1	Bruel & Kjaer 4003 omni-directional (x2) 1m from participants	4 channels at 48 KHz/24 Bit Using Audacity Phonic mixer console was used as a microphone preamplifier	JVC HD Everio GZ-HD7 high definition video cameras x2 Resolution: 1920x1080i Bitrate 26.6 Mbps
2	Beyerdynamic Opus 54 cardioid (x2) Head-mounted		

3.1.4 Evaluation of the material

There may be reasons to question whether the use of a recording studio is optimal when studying how speakers and hearers manage their interactions 'naturally'. That is, one is expecting individuals to behave naturally in a somewhat unnatural setting; i.e. (i) these people have perhaps never had a conversation in a recording studio before, and (ii) they do not start the conversation on their own initiative, but because someone has asked them to. For these reasons it is perhaps more preferable to collect conversational data elsewhere, in a more naturalistic setting. At the same time, there is nothing in the recordings suggesting that the participants conduct the conversation any differently from what conversations might look like elsewhere. They are still managing talk as it occurs naturally for them, in real time, and for these reasons the conversations, it could be argued, qualify as naturally-occurring conversations, and fit with the purposes of this thesis. The participants are somewhat constrained by the recording setting, but then there are also constraints in all naturalistic data, that interactants accommodate to.

A major advantage of the material collected in such a setting is the studio quality sound, and high-definition video quality. This is important for the kind of detailed phonetic and gestural analysis performed in this thesis.

3.2 Conversation Analysis

All the analyses in this thesis are data driven, which means that initial research questions were based on the observations made and developed while listening to/watching the recordings. In this way different phenomena were discovered that were then developed and explored using systematic analyses. The way this was done follows the traditional approach of Conversation Analysis (CA) for studying naturally-occurring talk-in-interaction (e.g. Heritage & Atkinson, 1984).

Two fundamental qualities of CA are (i) that it takes its analytic departure from the talk itself, putting the participants' own orientations in focus as they occur and develop in naturally-occurring interactions, and (ii) no level of detail is *a priori* regarded as unimportant to the interaction (Heritage, 1989). Generally, these qualities of CA motivate researchers to investigate and formulate new accounts for social action. According to Sacks (1992), "from close looking at the world you can find things that we couldn't, by imagination, assert were there...", and "if we can add to the stock of things that can be theorized about we will have done something more or less important – if the things that we've added have any import to them." (vol. 2, pp. 419-420). For a good example of how this approach can be put into practice, see Schegloff (1996a).

An important reason for using CA in this thesis is that it focuses attention on those structures that interactants themselves use and attend to during interaction. This can then be combined with a detailed investigation of phonetic and non-verbal resources. This section will in particular attend to the more structural components of CA, and present practical aspects regarding the use and implications of CA to this thesis. Subsection 3.2.1 focuses on the principle of finding and using evidence in a next turn, 3.2.2 will present some general consideration of sequential relevancies, and 3.2.3 addresses issues on quantification of CA research.

3.2.1 Next turn proof procedure

The most central aspect of CA is the attention to the interactional work that the interactants themselves do, and the categories that appear real to *them* in their management of talk-in-interaction. As far as possible, the validation of a certain analysis, or analytic category, comes from participants' own online orientations to the emerging talk. In CA, a key to validating/testing the interactional relevancies of what participants do is to use evidence in what happens next in the interaction, i.e. the so-called next turn proof evidence (e.g. Sacks, et al., 1974). In CA such evidence is used to establish for example how different linguistic features (e.g. lexical, grammatical, phonetic) affect meaning and the interactional progress.

To illustrate this point, I will use an excerpt and discussion taken from Schegloff (2007, p. 189), which is a telephone conversation between Ava and Bee. With this example, Schegloff (2007) addressed the interactional management sequence-closing, and of particular interest here is the interaction in lines 9-15.

```
(3.1) Schegloff, 2007, p. 189 "they have a problem"

 1      Bee:      There's only one time that I r-hh ˙hh- thet they really
 2                looked happy wz the time they were etchor hou(h)se.
 3      Ava:      Oh:. Yea:h. Didn' they look ha:ppy.=
 4      Bee:      =[Uhhh huhh! ˙hhh
 5      Ava:      =[Ho ho ho,
 6      Bee:      hhhunh [hunhh .hh
 7      Ava:            [Tha wz about ez happy ez they ge:t. Eh-hu:h,
 8      Bee:      ˙hh Really (now)=
 9      Ava:      =They have a prob'm.
10                (0.4)
11      Bee:      Mm.
12                (0.5)
13      Ava:       Definite pro:b'm,
14      Bee:      We:ll, ˙hh (0.3) I don'know.
15      Ava:       YOU HO:ME?
```

Schegloff (2007) makes two lines of arguments that both illustrate the relevance of next turn proof procedure. The first argument is that Ava's *they have a prob'm* in line 9 is an

example of the type of summary assessments commonly found when speakers are seeking to close a sequence/topic of talk. The second argument is that Bee's *mm* in line 11 indicates her reservations about agreeing with Ava. Regarding the second argument, the next turn proof lies in how Ava orients to Bee's *mm* as a reservation by upgrading her assessment in 13, with *definite prob'm* (i.e. the upgrading element is the use of the lexical item *definite*). Furthermore, Bee's more explicit expression of her reservations in 14 (*well ((...)) I don'know*) is evidence that Bee indeed had reservations with aligning with Ava in 11. Regarding the first argument, the next turn proof is found in how Ava initiates a new topic in line 15; this is evidence that Ava's assessment in 9 was indeed an initiation of a sequence closing.

3.2.2 Orientation to sequence

Another core point in CA is that talk is sequentially organised (Schegloff & Sacks, 1973; Schegloff, 2007), and the attention to the role and implications of such sequential constraints is central to the analysis of talk-in-interaction. With example 3.1 above, Schegloff (2007) argues that by providing a minimal *mm*, Bee violates the *preferred* structure in summarising/closing a topic, where typically (i) a speaker initiates a sequence-closing with an assessment, (ii) a co-participant aligns with this assessment (i.e. this is where Bee fails to provide a preferred response), and (iii) the main speaker ratifies this alignment, and a new topic can start.

Schegloff (2007) particularly refers to adjacency pairs, where a second turn (a Second Pair Part) is made conditionally relevant based on a first turn (a First Pair Part). That is, if a First Pair Part is produced and a Second Pair Part is absent, it will be noticeably absent. An example that clearly illustrates this point is given below (from Sacks, 1987, p. 64). In this example, A makes an agreement relevant in line 1, i.e. A produces a First Pair Part which in CA terms provides a preference for agreement (cf. Pomerantz, 1984). When an agreement is absent in line 2 (notice the pause), A proceeds to reformulate the First Pair Part into something more negative. This is not only observably responsive

to a lacking response, but also shows that A interprets the silence as a disagreement, 'in the making'.

```
(3.2) Sacks, 1987, p. 64 "good cook"

1    A:           They have a good cook there
2                 ((pause))
3    A:           Nothing special
4    B:           No, everybody takes their turn
```

Although sequential relevancies are central to the management of talk, the extent to which sequence plays a conditional role, and what constitutes such conditionality remains open to discussion (see e.g. Stivers & Rossano, 2010). This thesis seeks to elaborate some of the observations about structural constraints by attending to phonetic and non-verbal detail.

3.2.3 Quantitative and qualitative approach

In this thesis the arguments are made on the basis of a collection of instances, as in most of CA research. This is based on the assumption that actions (e.g. making a request) and sequences (e.g. adjacency pairs in request sequences) are comparable across instances. There are both quantitative and qualitative implications of this kind of research. In terms of the qualitative implications, CA research focusses on the significance of single case studies (e.g examples 3.1. and 3.2 above). At the same time, the collection and comparison of instances leaves room for quantitative analyses as well. In a majority of CA research however, the quantitative data are typically left unspecified. This might to some extent be based on the background of CA researchers (e.g. sociologists, linguists), but also on concerns regarding to what extent and for what reasons a quantitative analysis would benefit research of a given phenomenon. One question for example, is whether the frequency of e.g. overlaps between speakers, adds anything to the understanding of how overlaps in talk are managed, or what makes them relevant to the interaction (Schegloff, 1993). Still, there are no *a priori* reasons

why quantitative approach should not be used and specified in CA research (see e.g. Stivers, et al., 2009, for a study that combined CA-type questions with quantitative analyses)

In this thesis I will combine quantitative analyses with interactional/qualitative analyses (chapters 4 and 6). The quantitative analyses are meant to give an overview of the described phenomenon, providing a clearer indication of how common the phenomenon is, and in what circumstances it is found. The main significance in the analyses however, lies in a qualitative, single-case, approach.

A central part of the qualitative analyses is based on so-called *deviant cases*, examples that in one way or the other differs from a core set of examples, in that the talk takes a different route than what is expected. The rationale for paying attention to such deviant cases is that, although there are clear sequential constraints in how interaction develops, this is not deterministic. That is, sequential constraints are *resources* for interactants to use rather than *rules* to follow. By investigating how and why these examples are different from the core set of examples the aim is to get a clearer idea of what the central relevance of the phenomenon is, and also to improve the initial analysis.

As an example, Heritage and Sorjonen (1994) analysed the use of *and* to preface questions in institutional settings (e.g. "and what about…"). The role of these *and*-prefaced questions was described as providing a link between one question and a next one, and thereby giving the interaction a routine and agenda-based character. On this basis one might expect (i.e. as a rule) that the *and*-prefaced questions occur when a previous issue has been solved, allowing for the interactants to move on to a next one. However, as part of their analysis Heritage and Sorjonen (1994) provided some examples that deviated from this route. They found that similar *and*-prefaced questions were used following problematic issues raised in the interaction, or to avoid potentially problematic new topics. Thus, *and*-prefaced question can be used to 'normalise' problematic talk, and these deviant cases help enrich the initial analysis of this phenomenon. In this thesis I will support all of my analyses with the use of such deviant case procedures.

3.3 Data preparation and representation

In order to prepare the data for analysis, orthographic transcriptions were made (note that 'the data' here refers to the audio and video recordings, and not the transcriptions; these are representations of the data). Transcription is not only useful for representing data, it is also a good method for observing detail in the material. Thus, this was a natural part of the process in which research questions, and formulations about phenomena, were developed. The initial transcriptions were basic, with turns ordered in sequence and with rough annotations of overlaps and pauses. For the Call Home material a transcription was already available, however this was not of the detailed and sequential order that is preferable for CA, so for parts of this material I elaborated the already existing transcriptions. I then provided a more detailed transcription only for those parts that were of interest for my analytic purposes. For the transcription I used ordinary text editing software, employing ELAN[2] and Praat[3] as replay tools. ELAN is an audiovisual annotation tool important in particular for the gestural annotations in my thesis (introduced further in section 3.4).

3.3.1 Transcription conventions

When representing the data for an audience it is preferable to use a known/shared format, and the detailed transcriptions presented in this thesis are based on Gesprächsanalytisches Transkriptionssystem 2 (GAT2), developed by Selting, et al. (2009). This system shares most of the principles of the transcription developed by Gail Jefferson, which is the common transcription form within CA research (e.g. Heritage & Atkinson, 1984). In Gail Jefferson's conventions, words are freely transcribed according to their pronounced form: For example "with" can be seen transcribed as "wih", presumably because the transcriber has not observed any dental occlusion at the end of

---

[2] Free to download at http://www.lat-mpi.eu/tools/elan/
[3] Praat is a widely used software within phonetic research. It is free to download at http://www.fon.hum.uva.nl/praat/

the word. In my view, a transcription is more consistent, and also easier to read if it gives orthographic forms of words, rather than the mix of phonetic and orthographic forms found in Jeffersonian transcription (see Walker, 2004b, pp. 38-43, for a further discussion). This was one reason for using GAT2. Another reason was that GAT2 provides some conventions for representing prosodic features, which are more compatible with linguistic/phonetic representations, than other transcription conventions used in interactional research. In this thesis conventions for indicating prominence, intonation, speech rate, and loudness are used.

As has been noted in previous CA research (e.g. Walker, 2004b), an issue with including additional elements in the transcriptions is that it assigns analytic relevance to some phonetic/linguistic detail and not others. However, this is not seen as a major problem in this thesis. Prosodic information was included in the transcriptions to give the reader some further idea of how the talk was produced, even if the data are not accessible to them; the reader should in any case be aware that a transcription does not do full justice to the details of the data, and that he/she should consult the recordings to further access them.

The inclusion of prominence and intonation was done consistently throughout. Because transcription of other phonetic features, e.g. speech rate, loudness, and voice quality, might negatively affect the readability of the transcripts, these were only provided in cases where it was meant to support the analysis.

A summary of the transcription codes are given below:

Sequence of turns

[ ]	Overlaps between turns. Left bracket – start of overlap, right bracket – end of overlap
=	Latching, between the end of one turn and the beginning of a next, or connecting two lines that contain the same TCU

Breathing

°h / h°	In-breaths and out-breaths respectively, 0.2-0.5 sec
°hh / hh°	In-breaths and out-breaths respectively, 0.5-0.8 sec
°hhh / hhh°	In-breaths and out-breaths respectively, 0.8-1.0 sec

Pauses

(.)	Micro-pause, below 0.2 sec

(-)	Short pause, 0.2-0.5 sec
(--)	Medium pause, 0.5-0.8 sec
(---)	Longer pause, 0.8-1.0 sec
(1.0)	Longer pauses indicated by seconds

Durations

:	Prolongation of sound/syllable, 0.2-0.5 sec
::	Prolongation of sound/syllable, 0.5-0.8 sec
:::	Prolongation of sound/syllable, 0.8-1.0 sec

Accents/prominence

acCENT	Accented syllable in capital letters
ac´CENT	Rising pitch contour
ac`CENT	Falling pitch contour
ac¯CENT	Level pitch contour
acˇCENT	Falling-rising contour
acˆCENT	Rising-falling contour

Turn-final pitch movement

?	Rise to high
,	Rise to mid
-	Level
;	Fall to middle
.	Fall to low

Other conventions

?	Glottalisation
↑	Pitch step-up
↓	Pitch step-down
((head-move))	Non-verbal/non-spoken productions or events
(yes)	Candidate hearing
(he/you)	Possible candidates
<<p >word >	Describing loudness, speech rate and voice quality, and indicates where it starts (<< >) and ends (>). Codes: p – piano, pp – pianissimo, f – forte, ff - fortissimo, all – fast, lento – slow

The transcription of breathing, pauses and prolongations of speech sounds were done quantitatively (but are not strictly accurate) in the software. Other labels, e.g. pitch and prominence, were determined based on impressionistic listening, relative to surrounding syllables and speech elements. Prominence (capital letters) was assigned to those syllables in the turn that could be categorised as pitch accents in intonational analyses (see e.g. Cruttenden, 1997). The cues to pitch accents are normally based on a combination of duration, pitch and loudness. In Norwegian (and particularly East-

Norwegian, which is studied here), one correlate of pitch accent is falling or low pitch, relative to surrounding syllables (see e.g. Kristoffersen, 2000). This is different from for example English, where one correlate of pitch accent is rising or high pitch. Extra IPA symbols were used in connection to particular arguments regarding phonetic realisations.

An example (3.3) from the transcript of the Norwegian material will be used to further illustrate the conventions used.

```
(3.3) KTH-NO, AO, 07:50, "befinne seg"

01    O:           jeg har  til og  med  gått (i/eh) `TRE  år    på:
                   I   HAVE TO  AND WITH GONE (IN)    THREE YEARS ON
                   I have even gone three years for

02                 konversa`SJONskurser her  i [`STock]´holm,
                   CONVERSATION-COURSES HERE IN   name
                   conversation course here in Stockholm

03    A:                                      [mm;   ]
                                                mm

04    O:           og [det er] eh: (f)
                   AND IT  IS      (STILL)
                   and it's uh (s-)

05    A:             [mm,    ]
                       mm

06    O:           `FORT´satt `VANskelig å (.) [°h ] uttrykke seg
                   STILL      DIFFICULT TO          EXPRESS  refl.pron
                   still difficult to (.)       °h   express myself

07    A:                                      [ja,]
                                              YES
                                              yes

08    O:           [`FLYtend↑e] `altså; h°
                   FLUENTLY      THUS
                   fluently     you see

09    A:           [°hh       ]
                    °hh
```

All examples are headed with (i) transcript number (i.e. [chapter].[transcript]), (ii) corpus title, (iii) name of recording (for KTH-NO these are based on the initials of participants), (iv) time tag, and (v) name tag.

As far as possible, each speaker turn is assigned a line, numbered to the left. However, as is shown in example 3.3, a speaker turn sometimes extends more than one line. The relevant speaker is indicated to the left, following the line number. If there is no name-initial, this means that the speaker from the previous line continues (cf. line 02).

Each line in the transcription is separated with a paragraph, and in each line the Norwegian transcription is given first, followed by a translation gloss (capital letters) and a pragmatic translation (italics) to English. The translation gloss is carefully aligned with the associated word in Norwegian, but this is not done for the pragmatic translation. The translation gloss gives a word-by-word translation, with morpho-syntactic elements embedded in the translation: e.g. "språk" (singular) is glossed as LANGUAGE, whereas "språk" (plural) is glossed as LANGUAGES. In cases where a word in Norwegian does not have a direct translation in English, grammatical tags in lower case letters were used. See for example "seg" – 'refl.pron' (reflexive pronoun) in line 06 (a further list of such tags is given in Appendix B). Further, a potential translation is given in brackets '( )' when a word is not fully produced (see (STILL) in 04).

3.4 Micro-analytic analyses of non-verbal behaviour

This section presents some fundamental procedures leading to the analysis of non-verbal elements in this thesis (particularly chapters 5 and 6), and how these will be presented, as an addition to the transcription conventions described above. The term 'micro-analysis' is suitable to describe this work. Micro-analysis has been used by Loehr (2007), who determined, and quantified, the timing relation between verbal and a range of non-verbal elements (cf. chapter 2). In this thesis, micro-analysis is conceptualised as an additional component to CA, providing more detail to speech production than CA usually does. The micro-analytic work in this thesis relates first and foremost to the use of manual gestures, but also to the use of head-nods and gaze, along with speech. First I will present the tool (ELAN) used to conduct these analyses (3.4.1). Then I will provide basic conventions for segmenting and labelling non-verbal

elements (3.4.2), followed by a description of how non-verbal details are represented in the thesis (3.4.3). These descriptions are only relevant for the Norwegian material.

3.4.1 Annotation of audiovisual data

In all of the analyses, video-analysis and observational methods are used to make decisions about the relations between speech and non-verbal conduct, i.e. no quantitative means were used (e.g. motion detectors or other technical equipment). The analyses were performed in ELAN. ELAN makes it possible to perform combined audiovisual analysis, as one may create a simultaneous output of audio and video files. For my material, the raw data was an audio recording (stereo) and two video recordings. These were brought into ELAN and then synchronised manually. This process was eased by the fact that a clapperboard was used in the recordings, to synchronise picture and sound.

ELAN is based on a tier system which makes it possible to annotate the data on multiple tiers. I used this tier system to annotate verbal and non-verbal conduct for each speaker. The main purpose of this annotation was to determine the timing relations between verbal and non-verbal conduct, which then would be used as part of the interactional analysis. Timing relations were not quantified as in Loehr's (2007) work. Figure 3.B is a screenshot from an ELAN file.
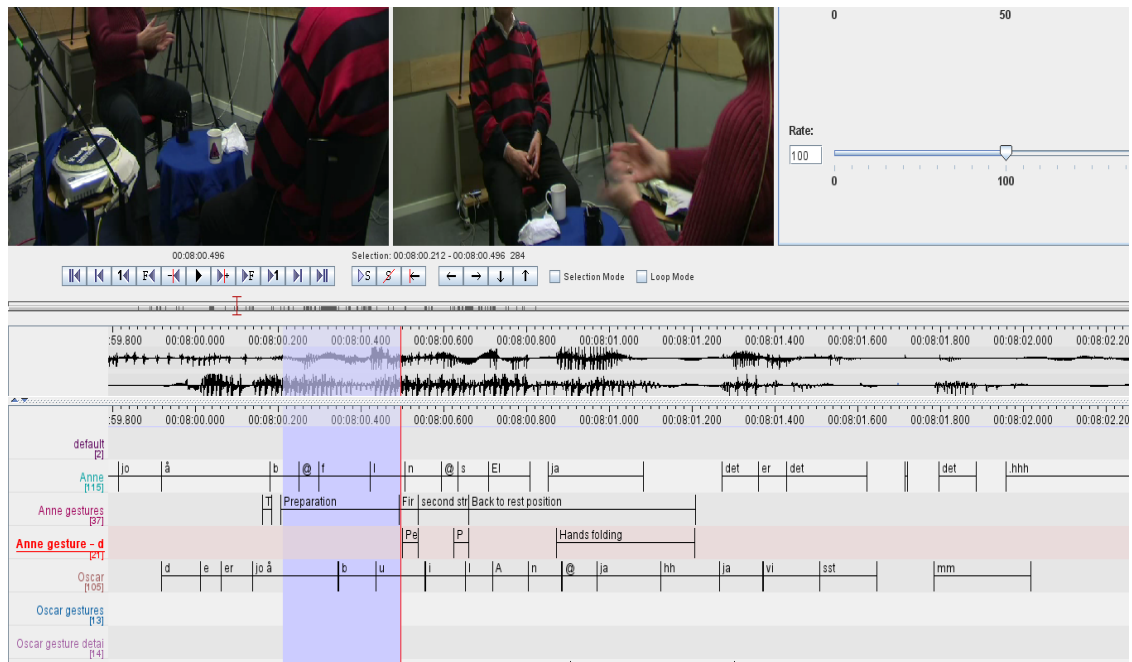
Figure 3.B Excerpt from an ELAN project, showing two synchronised videos (top), audio (waveforms in the middle), and labelling tiers (bottom).

To determine the alignment of non-verbal elements with speech, verbal production was labelled according to segment. In general, verbal and non-verbal elements were annotated separately, which for non-verbal annotations meant that the sound was turned off. This was done to be sure that the annotation was not affected by spoken productions.

3.4.2 Segmentation and labelling of non-verbal behaviour

The non-verbal behaviour described in this thesis is mainly head-nods, manual gesture and gaze, and the conventions for defining and labelling these will be described next.

Head-nods. Head-nods are defined as vertical or left-to-right/right-to-left movements of the head that involve movements that take at least one two-step motion (i.e. up and down, or left and right), and that are continuous over time rather than discrete. Judgments regarding head-nods were based on whether or not there was any such observable movement. There was no explicit lower limit for what constituted a head-

nod, apart from being sure to have identified movement, based on the above definitions. Head-nods were not segmented into constituent parts, e.g. whether the head is currently pointing upwards or downwards, they were labelled only according to whether or not there was one. However, some of the analysis makes relevant the distinction between 'regular' and more intensive nodding.

Manual gestures. Manual gestures were defined as movements by one or both hands, which played some kind of representational, pragmatic or interactional role in the emerging talk (see e.g. Kendon, 2004, for further background on classifying gestures). As a means of accurately and consistently determining the timings of gesture with speech, gestures were segmented into constituent parts, based on the definitions by Kendon (1972; 2004). The most important elements for the annotations performed in this thesis are: The preparation stage, the stroke, and the release of gesture, i.e. what constitutes a 'gesture unit' in Kendon's (2004) terms.

The stroke is the main part of a gesture, and in this thesis the stroke is defined as the part of the gesture where the handshape aimed at is ready and the hand moves in the direction of the peak of the stroke. The stroke peaks were labelled and defined as the physical end-point of a stroke. Preparation is defined as the initiating part of that movement and handshape. Apart from these categories, it was determined whether a stroke would be held following its peak, and at what point the stroke/hold would be released into resting position (i.e. no gesture), or reshaped into a new gesture.

As mentioned, the purpose of this segmentation was to give a clear and consistent description of the development of gesture with speech. Note that it is not *a priori* assumed that these gestural segments are meaningful in terms of perception, i.e. that they *have to* be done in order to analyse gesture adequately. Rather, by providing such segments it was assumed that these categories and their boundaries would form the basis for analysing how the timing of gestural events matter for the interactional process.

Gaze. Although gaze is not a major topic in this thesis, it will frequently be referred to and used as part of the analyses. As is shown by previous work (e.g. Kendon, 1977, 1990; Goodwin, 1981; Hayashi, 2003b), gaze is a powerful resource used in the

organisation of turns and in creating participation frameworks. In this thesis gaze was labelled according to whether or not there was mutual gaze between the speakers, and if not, the direction of the gaze. This was determined based on observational measures only. Due to the high quality of the videos, it is reasonable to assume that the gaze labelling was accurate.

3.4.3 Transcription conventions for non-verbal behaviour

Annotations for non-verbal behaviour are placed above the verbal transcriptions in the examples. The convention for representing head-nods and manual gestures on paper are loosely based on those of Kendon (2004), whereas transcription of gaze is inspired by the conventions used by Goodwin (e.g. Goodwin, 1981).

For transcriptions of gaze I use the following categories and symbols:

___	(continued) gaze at co-participant
x	point in time where mutual gaze is achieved
, ,	gaze away from or towards co-participant
U	gaze direction: up
D	gaze direction: gaze down
R	gaze direction: gaze right
L	gaze direction: gaze left
+	eye-blink (only transcribed where relevant for analysis)
e.g. DR	gaze direction: down right (DR), up left (UL), etc.
{table}	specifying object gazed at

And the conventions for representing head-nods in the transcriptions are:

^^^	Vertical nodding
<><>	Left-to-right nodding
^v^v	More intensive (vertical) nodding
//	Start/end of a nodding unit
/	Dividing subcomponents of head-nods, e.g. changing in intensity

In chapter 5 transcriptions of gaze and head-nods will be combined, and example 3.4a shows what that looks like. Gaze ('Gz') is given in green and head-nods ('HN') is given in brown. Here we can see that Lars gazes at Bengt as he produces *sett*/"seen" (02), and Bengt starts nodding during his first *mm* (03-05), while Lars gazes at him and continues speaking (04-06).

```
(3.4a) KTH-NO, BL, 04:20, "aleine" HEAD-NOD AND GAZE ANNOTATION

02     Gz(L)                      , , , , , x_____
02     L:          jeg ikke skulle ha ‾SEtt:— (0.3) om je:g— (eh)
                   I   wouldn't have   seen: (0.3) if I: (uh)

03-06  Gz(L)          _____, , , DR
03-06  HN(B)       //^^^^^^^^^^^^^^^^^^^^^^^^^^^//
03-05  B:           mm, =                        [mm;    ]

04-06  L:              = hadde gått a`LEIne? [for ekse]mpel: (eh)
                         had gone on my own    for example (uh)
```

The initial in parenthesis indicates who the producer is. Note that for gaze, normally only speaker's gaze direction is specified. This is because hearer normally gazes at speaker (cf. Kendon, 1977). Hearer's gaze is then only specified when it deviates from this norm (e.g. hearer gazes away from speaker), and this will be indicated in a separate line. The placement of the non-verbal symbols are meant to align with the verbal productions, i.e. Lars starts moving his gaze towards Bengt during the onset of *skulle*/"should" in line 02. Notice that the symbol '=' is used to show where the transition between speakers happen; in this example Lars' *hadde gått aleine*/"had gone on my own" starts immediately after Bengt's *mm*.

Notice also that the line number is specified to the left. In many cases, this is more than one line (e.g. gaze and head-nod transcription in 03-06). The numbers refer to the line numbers in the main transcription, which is based on verbal elements only, and organised according to conventions about what constitutes a potential turn completion (see chapter 2, section 2.2). In most cases, the main (verbal) transcriptions are presented first, followed by the verbal + non-verbal transcriptions. This is done to make it easier for the reader to access the data. In this case, transcript 3.4a above refers to transcript 3.4b below. Arrows indicate which turns are presented in transcript 3.4a.

```
(3.4b) KTH-NO, BL, 04:20, "aleine"

01    L:              °pthh jo  `DET ↑var `KUlt.
                           (YES) THAT WAS  COOL
                      °pthh yeah that was good

02    L:              og  jeg fikk jo  `SE  saker  som
                      AND I   GOT  part SEE THINGS THAT
                      and I   got to see  things that

              ->      jeg ikke skulle ha   ⁻SEtt:⁻ (-) om je:g⁻ (eh)=
                      I   NOT  SHOULD HAVE  SEEN       IF I
                      I wouldn't have seen:      (-) if I: (uh)

03    B:      ->      =mm,
                       mm

04    L:      ->      hadde gått a`LEIne?
                      HAD   GONE ALONE
                      had gone on my own

05    B:      ->                  [mm;    ]
                                   mm

06    L:      ->      <<all >[for ekse]mp>el: (eh) om du  har ´HØrt  om
                             FOR EXAMPLE          IF YOU HAVE HEARD ABOUT
                             for example (uh) have you heard about those
```

The purpose of including several lines from the original transcription into one in the non-verbal the transcriptions, is to give a more continuous representation of the co-ordination between non-verbal, verbal and inter-speaker behaviour. Also, when adding non-verbal information it becomes less straightforward to represent the data in a discrete, line-by-line fashion, than when transcribing verbal elements only.

As with head-nods and gaze, manual gestures are represented on top of the verbal elements. The transcription symbols for manual gestures are as follows:

....   Movement of hands - preparation for stroke or withdrawal

^   Stroke of gesture

x   Peak of stroke

---   Gesture hold

//   Start/end of gesture unit

/   Separating elements within a gesture unit

( )   A weaker tendency of gestural movement, stroke, peak or gesture hold

An example of transcriptions with manual gestures ('MG') is given in example 3.5 below. In this example the reader will notice that still-shots from the video recordings are included, showing the relevant manual gesture. The three still-shots represent the preparation (figure a), the peak of the gesture stroke and its hold (figure b), and the release of the gesture (figure c). The exact placement of these stills-shots in the emerging talk is indicated with a line pointing towards the transcription.

```
(3.5) KTH-NO, TL, 7:13/552 "Torbjørn Thorsen" GESTURE ANNOTATION

01    T:           Torbjørn ´THOR`sen.
```

a   b   c

```
02-04  MG(T)             //....^^x--------------------........//
02-03  L:          [(-)[(-) [Torbjørn `THOR´sen ja; ´HA[N kjenner jeg `go]dt.
                            Torbjørn Thorsen (yes) I know him well

02-04  T:          [(-)[(-) [                         [mm,
```

Still-shots are only provided for annotations of manual gestures, and not for head-nods and gaze. This is to protect the participants' identities.

Transcription principles similar to those described above will on occasions also be used when describing facial movement and other non-verbal behaviour.

3.5 Summary

This chapter has attended to practical issues relevant for all, or at least two, of the analysis chapters. It introduced the material collected and used in this thesis, the main method for exploring the data (CA), and the conventions and approaches to analysing the data and representing the data on paper. More particular aspects of the analysis will be presented in each analysis chapter.

CHAPTER 4

PHONETIC RESOURCES FOR 'DOING THE SAME'

One important task for hearers is to, in real conversational time, show whether they intend to end their role as hearer and project a next speakership, or continue being hearers. This chapter investigates how hearers, using phonetic resources, maintain and differentiate their actions during a speaker's turn, and how this affects the negotiation of speaker change, and whether or not talk continues on the same topic.

As an illustrative starting point for the study, example 4.1 is presented below (from the Call Home corpus). This example shows how a hearer may actively disengage from a current topic, by producing two similar and minimal responses to two linked but separate elements in a speaker's turn. At this stage, particular attention will be given to what 'similar' means in action terms. This will then be the focus of the combined interactional and phonetic analysis that follows.

In lines 02 and 05 of the transcript, Gerry (Ger) provides a negative assessment about Lisa Marie Presley, who Gerry and Patricia (Pat) independently have seen on a televised interview during the period when she was married to Michael Jackson. Rather than explicitly agreeing with Gerry, Patricia seems more willing to talk about another part of the interview (line 08). Prior to that topic-shift, Patricia produces *yeah* twice in response to Gerry's assessment, providing only a minimal alignment with Gerry. The two *yeah*s are marked with arrows in lines 03 and 06.

```
(4.1) CH4092, 1613-1620 "Lisa Marie Presley"
01    GER:         (°ph) but I just ^SAW it fo? I just ´SAW like maybe the
                   ´LAST ten minutes-
02    GER:         °hh but ´SHE's such an ´IDi`ot.
03    PAT:  ->     yeah(m).
04                 (.)
05    GER:         it's ^TRU:ly ¯UNbelievable how ^S:TUpid she is.
06    PAT:  ->     yeah(m).
07                 (--) /GER: (°hh) /PAT: (°p)
08    PAT:         °thh well now did you ´SEE the ˘PART [where    ]
09    GER:                                             [I ´GUESS ] her
                   ´F:A`ther ´rea`lly wasn't that bright either but-=
10    PAT:         =no I don't think so;
11    PAT:         [°hh]
12    GER:         [mm ] anyway okay;=
13    PAT:         =but did you ´SEE the ˘PART whe:re ((---))
```

Assessments generally make relevant some sort of agreement from a co-participant (Pomerantz, 1984), and by lexically upgrading her second assessment in 05 (i.e. with *truly unbelievable*), Gerry gives Patricia a second opportunity to agree with her. But instead of explicitly orienting to such an opportunity, Patricia produces a second *yeah* in 06, followed by a shift in topic (08). Patricia's two *yeah* responses are clearly minimal in terms of the sequential relevancies here, and it seems like they are designed as minimal in order for Patricia to project her topic-shift.

As supporting evidence for Patricia's willingness to shift topic, is that her topic-shift is continuing on a telling she initiated previously, about something in that interview that she and her colleagues had been laughing at at work[4]. Furthermore, evidence that Patricia's responses are oriented to as minimal, and to the topic-shift as inappropriate, is found in how Gerry continues on her own topic (09) in overlap with Patricia's new topic (this time comparing Lisa Marie with other members of her family). This shows that Gerry does not treat her own actions in 01-05 as adequately accomplished, and thereby treats Patricia's shift as sequentially unfitted. After having secured an explicit agreement from Patricia (10), Gerry then gives the next turn, and the rights to continue on the projected topic-shift, back to Pat (12)).

In sum, Patricia's two *yeah*s are evidently taking part in negotiating a topic-shift. Now, the important questions that follow in connection to this are:

- Is there something in the phonetic production of Patricia's second *yeah* (06) compared to the first *yeah* (03) that indicates her disengagement with Gerry's talk?

- Are such phonetic characteristics used systematically (i.e. across instances), as distinctive from phonetic characteristics of engagement with current talk? And do these phonetic characteristics have similar effects across lexical categories?

- Are these differences oriented to by the participants when they negotiate towards a next turn?

---

[4] Excerpt of interaction prior to and following the excerpt in transcript 5.1. Patricia initiates the telling in lines 01-02, and summarises it in line 41 (arrowed lines).

```
01    PAT:   ->      and ´Ever since that ´INterview `like;
02           ->      °h everyone ^QUOTes Lisa Marie Presley at ´WORK,
03                   HH°:: heh
04                   (-)
05    GER:          (p?) °h she sai? she (just) was `SO `DUMB in that interview:.
06                   I mean the [´BIT that ˇI sa]w:,
07    PAT:                      [oh you ˇSAW it?]
              ((3 lines omitted))
11(01) GER:          (°ph) but I just ^SAW it fo? I just ´SAW like maybe the ´LAST
                     ten minutes-
              ((18 lines omitted; see transcript 5.1))
30    PAT:           °h and lisa marie goes. (--) °pthh ´I'D tell
                     them to ´EAT `me; HHhh° hah hah hah hah hah=
31    GER:           =she 'SAID ´that?
              ((10 lines omitted))
41    PAT:   ->      so ´NOW we do that at work <<laughter> all the ˇTIME?>
```

These are the main questions that this chapter seeks to answer, by conducting phonetic analyses of hearer responses combined with sequential-interactional analysis, mainly in Norwegian conversation. As an interactional study, the main objective of this analysis will be to show one way in which hearers may take an active role in projecting a shift in talk, while attending to ongoing talk. But this study also deals with issues related to linguistic-phonetic variation and variability. That is, it seeks to show how verbal responses, with their phonetic and lexical characteristics, need to be understood as *part of* a sequential-interactional environment, and that by studying such relationships one may account for some phonetic variability not previously accounted for.

These issues will be addressed further in the background section (4.1), along with a consideration of how this particular study relates to previous research on hearer action with the use of response tokens. Following the background, the procedures and analysis for the study will be presented in three steps. First, in section 4.2, the procedures for collecting comparable instances and defining the relevant action categories will be described (4.2.1), supported by interactional evidence in a set of examples (4.2.2), and their distributions across action categories (4.3.3). This will form the basis for the phonetic analyses presented in section 4.3. Section 4.4 will provide further interactional evidence for the main findings, using two examples that deviate from the core set of examples. Finally, section 4.5 will offer a summary and discussion.

4.1 Background

The main target of the analysis will be hearer actions in a particular sequential-interactional context. The hearer actions investigated in this study involve the use of verbal responses like "yeah" and "mhm" (English), which are frequently referred to as 'back-channels' in the non-CA literature (e.g. Yngve, 1970; Duncan, 1974), and have been referred to as 'continuers' (Schegloff, 1982) 'acknowledgment tokens' (Jefferson 1985, 1993; Drummond & Hopper 1993; Gardner 2001), 'reactive tokens' (Clancy, et al., 1996) and 'verbal feedback' (Stubbe, 1998), in the CA literature. The point of the CA categories seems to be to give the relevant response tokens more action-oriented

names, as they take active part in the developing interaction, and not simply maintaining the on-going talk (i.e. 'back-channels'). In this chapter, 'hearer response tokens' (and variations thereof) is generally used, except when reporting findings from studies which have explicitly used specific terms, such as 'back-channels'. This term might be less specific in terms of action than some of those mentioned above. However, this seemed a natural choice, as the study focuses on the interactional process which involves these responses, rather than on (potential) individual functions/meanings of these response tokens.

As a relevant background for this study, previous studies on response tokens will be reviewed, particularly in terms of the relation between phonetic characteristics and functional/interactional meaning. In section 4.1.1 I will show why it is important to base a study of response tokens on their sequential placement and development, rather than on pre-established categories of meaning or function. Then in 4.1.2, previous research on how response tokens work in a sequence, and in relation to a (potential) speaker change, will be presented. Although response tokens have been investigated in phonetic terms as single responses (e.g. Ward & Tsukahara, 2000; Gardner, 2001; Benus, Gravano, & Hirschberg, 2007), and in sequential terms as a chain of responses (Jefferson 1985, 1993), no known study explores the interrelationship between sequence and phonetics specifically for a chain of response tokens, as this study does. However, the interrelationship between sequence and phonetics has been addressed regarding lexical repetition in talk (Couper-Kuhlen, 1996; Curl, 2005; Curl, et al., 2006), and 4.1.3 will pay particular attention to one of these studies.

4.1.1 The 'meaning' of response tokens based on phonetics

The functional complexity of response tokens is acknowledged by several researchers. Based on the range and variability in terms of their lexical and phonetic production, response tokens have been described as highly ambiguous items (Stubbe, 1998; Benus, et al., 2007), far from being adequately understood (Gardner, 2001). This is not surprising given the different actions response tokensmay perform, and the different

sequential circumstances they occur in. Nevertheless, there have been few serious attempts at approaching these complexities, in accounting for phonetic (and lexical) variability.

Typically, in previous studies, categories of response tokens are assigned based on the analysts' native understanding of what the pragmatic function of individual responses might be. For example, in addressing the phonetic characteristics of (English) response tokens, Benus, et al. (2007) categorised different types of hearer responses according to discourse functions, including (continuer-type) back-channels, agreement and affirmative response types. They found some lexical and prosodic regularity based on these categories: In terms of pitch contour, back-channels consistently had steeper pitch slope (i.e. final rising pitch) than other types of responses.

This study suggests that the functions of response tokens are clearly associated with their phonetic characteristics. However, it remains unclear whether and how these phonetic characteristics are indeed relevant for the participating speakers and hearers themselves, and for the interactional process. Also, these findings assume that the phonetic form of a response token determines function, and that these functions are stable, irrespective of where the responses occur in talk. Such an approach seems to account for meaning as a bottom-up process, rather than as a top-down process, which would involve the interactants' knowledge of what occurs when in a particular structure, e.g. in a particular sequence of talk (see also Couper-Kuhlen & Selting, 1996, for a further criticism of decontextualisation of linguistic/intonational forms; and Goodwin, 1986, for a critique on back-channel research).

When considering contextual information, previous studies commonly attend to the prosody of the preceding speaker's turn (e.g. Ward & Tsukahara, 2000), but rarely attend to the interactional context. Gardner (2001) is an exception. Based on interactional evidence he suggested that the functions of "mm" responses depend on their pitch contour, distinguishing (i) a continuer-type response (rising/flat contour), (ii) a 'weak', somewhat disengaging, acknowledgment (falling contour), and (iii) a more affirming acknowledgment (rising-falling contour). But although Gardner (2001) introduced rich interactional detail into his analysis, linguistic form is still to some extent

treated as decontextualised, i.e. it is implied that the meaning of response tokens lies in their prosodic forms, and thus for example "mm"s with a specific prosodic pattern carry a general meaning.

Instead of limiting the connection between form and function to a (single) response token, this study pursues the possibility that also the phonetic relationship between consecutive responses may, in certain sequential environments, be consequential to the interaction, and that this variability might work both within and across lexical categories. Before reviewing the current study and its potential implications in relation to previous work, it is worth revisiting some fundamental aspects of the current approach (see also chapter 2). In this study, the basis for investigating how response tokens acquire their functions is to discover systematic ways in which interactants draw on, and combine, phonetic and lexical characteristics to distinguish their actions, in orientation to the interactional-sequential process. In doing so the interactants use previous knowledge of structure *and* information from the incoming signal to determine what is going on in the interaction, and this study tries to pin down how phonetic detail is used as part of this process, by controlling for interactional-sequential context. That is, the main question is whether phonetic detail can be used systematically to distinguish interactional 'options' (or choices) that are already relevant on the basis of context. In this way the current study, and thesis, does not treat bottom-up and top-down processes as necessarily separate, and does not 'choose' one approach in favour of another, but views them as parts of the same process.

4.1.2 A sequence of response tokens and the projection of a next turn

A sequential account of hearer responses has been provided by Jefferson (1985, 1993), who noted that hearers regularly move from "mhm" to "yeah" (in English) when they intend to speak next. These findings were also replicated by Drummond and Hopper (1993). An example from Jefferson (1985, p. 7) is given below. Note how hearer E first responds with a *Mm:hm* in line 11, followed by a *yeah* in line 15, at which point E also projects a next turn.

```
(4.2) Transcript from Jefferson (1985, p. 7)


  (1.1.a) [NB:I:6:R:9:Standard Orthography]
  1    L:        I didn't have five minutes yesterday.
  2    E:        I don't know how you do i:t.
  3                    (0.3)
  4    L:        I don't kno:w. nh hnh
  5    E:        You wuh: work all day toda:y.
  6                    (0.3)
  7    L:        Ye:ah.
  8                    (0.2)
  9    L:        Just get Well I'm (.) by myself I'm kind of cleaning up
  10             from yesterday.
  11   E:   →    Mm: hm,
  12                    (0.2)
  13   E:        ˙t˙hhh hhh
  14   L:                 [°A-and° (.) °I was just g-washing the dishes,°
  15   E:   →    Yeah we're just (.) cleaning up here too:.
```

This shows a differentiated use of lexically different response tokens, to distinguish between hearership and projected speakership. As to whether certain phonetic characteristics (in similar lexical items) might lead to similar distinctions is unclear. The analyses by Jefferson (1985, 1993) and Drummond and Hopper (1993) seem to favour the conclusion that hearers distinguish their actions mainly on the basis of lexical distinctions, and maintain their actions with lexically similar tokens, irrespective of their phonetic characteristics. One exception is Jefferson (1993), who provided some examples where a "yeah" is shaped differently from a preceding "yeah", and she suggests that this distinction is relevant in relation to the projection of a next turn; in

these cases specifically for the initiation of a new topic. One of these examples is presented below (Jefferson 1993, p. 5). Jefferson (1993) describe the *yah*s in lines 5 and 10 as having a flatter intonation than the *yah* in line 4, and Jefferson suggests that the latter two are the least topically engaged. K follows on these 'flat' *yah*s with a topic-shift (line 12), in overlap with C's continuation.

(4.3) Transcript from Jefferson (1993, p. 5)

```
(1.4) [SBL:2:2:3:R:5a:Standard Orthography]

 1  C:         Well tha:t's one rea:son I don't want three tables
 2             up here Kate 'cause: because'hhhh my: hou:se is just
 3             to[o:  s:]ma::ll=
 4  K:           [Ya:h.
 5  K:        =[Yah.       ]
 6  C:         [A:ND uh't'hh if I ha:ve a(w) another wuh In the
 7             frontroom and in the dining room's fi:ne but if I
 8             ha:ve one in the KItchen over he:re.
 9             (.)
10  K:  →      Yah.=
11  C:         =hhhh[hh]Well then i]t's too ↑clo::se[see,        ]
12  K:  →           [Di]d you neh-                  [Did y]ou
13             notice: uh Claire I think you came over and played.
14             the time I had three tables with these all these
15             other gals,
16             (1.0)
17  C:         eYe::a:h?
```

What becomes apparent in this example is that phonetic detail is one resource with which hearers may distinguish and maintain their actions. However, neither Jefferson (1985/1993) nor Drummond and Hopper (1993) provided any systematic or detailed phonetic analyses in their studies. That will be one main contribution of the current study.

4.1.3 The interrelation between sequence and phonetic detail

Although not specifically related to response tokens, there are some interesting findings related to the role sequence plays in relation to the phonetics of lexical/syntactic

repetition in talk-in-interaction (e.g. Couper-Kuhlen, 1996; Curl, 2005; Curl, et al., 2006). Curl (2005) showed how the repetition of lexical/syntactic form can deal with one kind of interactional relevance, whereas the phonetic/prosodic form of the repetition, can deal with another. The lexical/syntactic repetition in focus occurred as the co-participant displayed a possible problem in hearing (i.e. initiation of repair). It was demonstrated that while the lexical/syntactic construction stayed the same, the phonetic characteristics of the repeat (the repair) depended on whether or not the repeated (the repaired) turn was fitted (as opposed to disjunctive) with the co-participant's talk. Importantly, Curl (2005) shows that there is not necessarily a one-to-one relationship between linguistic form and meaning, and that the phonetic production of a turn of talk is sensitive to the sequence in which it occurs. This has clear implications for how we understand lexical (and grammatical) meaning, which is something I intend to pursue in the current study, for response tokens.

4.1.4 Summary

This background has shown that response tokens are typically studied as single items, either based on analyst's perceptions and categories of functional role (e.g. Benus, et al, 2007), or based on interactional evidence using CA as a method (Gardner, 2001). Few studies investigate a chain of hearer response, and those that do, focus on lexical distinctions, and not on how (non-lexical) phonetic characteristics may add to, or work independently of, lexical distinctions. This calls for a study that takes seriously both the sequential, interactional and phonetic aspects of hearer actions in the form of verbal responses.

4.2 Interactional analysis

This section focuses on the hearer *action*, and what it means in interactional terms for hearers to maintain and differentiate their actions, while producing response tokens. The interactional analysis will formulate a distinction between two action categories, in a particular interactional context. In this interactional context, what hearers do when maintaining their actions across responses will be referred to as 'doing the same' in this study, as opposed to 'NOT doing the same', which refers to hearers' differentiation of action. These action categories will be used as a basis for the phonetic analyses in section 4.3, investigating the phonetic realisations of the two action categories. Thus, this section formulates the interactional 'control' for studying the relevance of phonetic detail in a particular interactional context.

Subsection 4.2.1 describes the procedures behind the definition of interactional context, and action categories 'doing the same' and 'NOT doing the same'. Illustrative examples for the two action categories will then be presented (4.2.2), followed by findings regarding how the action categories distribute according to what happens in the next turn, and according to lexical categories (4.2.3).

4.2.1 Procedures and definitions

As part of the analytic process, a sequence of interest was defined. This particular sequence is a relatively open-ended one, meaning that the transition from one speaker to another is from the outset negotiable rather than made *conditionally relevant* by the first turn. An example of the latter is following what Schegloff (2007) refers to as a First Pair Part (cf. chapter 3, section 3.2.2). Following a question for example, it is quite clear who should be speaking next, and the interaction might not progress until a fitted answer has been provided.

Other examples of a sequence type that makes relevant next talk by a particular speaker are so-called pre-sequences to tellings (e.g. Terasaki, 1976). Basically, a hearer's job in such cases is to provide a go-ahead for the telling to continue, i.e. *not* to take the

next turn. In contrast, in the sequence studied here next speakership is more of an open opportunity for a hearer, which may require that they take a more active role in distinguishing whether or not they project an uptake on the current talk. The turns prior to the hearer response may make relevant for example agreement, affiliation or appreciation, which hearers may or may not provide an explicit uptake on. Example 4.4 below is an (English) example, representing the kind of structure examined in this study.

```
(4.4) CH5736, 117-159 "Santa Cruz"
((Traci and Jill are discussing where Traci and her family could find a future home.
Traci wants to get a job at a college, at the same time as she wants to live close to
country-side surroundings. In 01 Jill considers the problems/challenges involved with
commuting))

01    JIL:          if you want ^HILLS ^AND a ^COLLege you're gonna have to (.)

                    ´TRAvel—

02    TRA:          (°ptk) yeah;=

03    JIL:          =every day—

04                  (--)

05    TRA:          =°ptk yeah; mh°
```

Notice that in 4.4 there are two opportunities for Traci to project an uptake, in lines 02 and 05, first following a complete TCU in 01, then following an add-on to that TCU in 03. Such add-ons are referred to as 'increments' in the CA literature (e.g. Schegloff, 1996b; Walker, 2004a; Auer, 2007), defined as a unit of speech that works as a syntactic constituent and a (in action terms) continuation of a preceding turn. All examples in the current study have in common that hearer responses occur on either side of an increment, however the definition of increment is somewhat wider in the current study compared to previous studies.

The features common to the sequence of interest are summarised below (Table 4.A).

Table 4.A. The sequence of interest for the study, shown and described in a turn-by-turn manner.

Turn unit	Speaker	Turn description
1	A (main)	*Turn in progress, coming to a possible (TCU) completion (the host). Constitutes a potential sequence-closing*
2	B (hearer)	*Response token #1*
3	A	*'Add-on'/increment to previous turn/TCU: Shaped as being part of A's previous turn unit, in terms of syntax/action*
4	B	*Response token #2*
5	A/B	*Who speaks next and about what depends on whether or not hearer is 'doing the same', displayed phonetically in the relation between response #1 and #2*

The main criterion for turn unit 1 (the host) is that it ends with a syntactic completion, and is complete in action terms, i.e. it constitutes a complete TCU. This includes cases where turn unit 1 ends with a conjunction (e.g. *og*/"and", *men*/"but" in Norwegian/English): Such conjunctions were considered part of a (possibly) complete turn unit if the conjunction was designed as part of the same intonation phrase as preceding turn elements, and was otherwise (e.g. in terms of speech rate and loudness) not heard as initiating a next intonation phrase, or projecting further speech using e.g. glottal stop as turn-holding cues (cf. Local & Kelly, 1986) (see also Jefferson, 1983, on how some turn-final conjunctions do not clearly project a continuation, and are vulnerable to overlapping talk). In addition to conjunctions, some anaphoric expressions (*det*/"it/that") were used in a corresponding way in the Norwegian material.

In all the cases the increment in turn unit 3 occurred as hearer did not produce (or project any uptake at his/her previous opportunity to do so. There may be several different interactional categories of increments to be accounted for (cf. Walker, 2004a; Auer, 2007), however for this study increments were defined broadly, as a continuation of syntax and action of a possibly complete turn/TCU construction, and not in themselves constituting a complete TCU. Again, the idea is that the host-increment relation remains to some extent constant throughout the examples, while the extent to which hearers differentiate their responding actions varies, having interactional consequences for the next turn, turn unit 5.

Regarding hearer response, cases were collected if response 1 and 2 was analysable as responding to turn unit 1 and turn unit 3, respectively. In the main analysis there were no restrictions as to what the response's lexical category would be, as long as they were single verbal items and functioned as some sort of acknowledgment. This included items commonly regarded as lexical (e.g. "yes" and "no"), but also items like "mm", which by some researchers might be thought of as less specified in terms of lexical meaning than a "yes". However, as this study will show, "yes" and "mm" can be used to do similar kinds of actions, and I find no straightforward reason to treat one as lexical and the other as non-lexical (see also discussion in section 4.5).The responses that occurred in overlap with either the end of the host or increment were included for the current purposes. Cases of head-nods were labelled, but will not receive major attention in the current study.

Examples were labelled as 'doing the same' ('dts') when there was no interactional evidence that a hearer differentiated their actions in response to turn unit 3 compared to turn unit 1. In contrast, examples were labelled 'NOT doing the same' ('Ndts') when there was such evidence. Interactional evidence was in particular based on what would happen next in terms of uptake on current topic. For 'dts', there would be no evidence of projected uptake from hearer on current talk, meaning that either speaker B does not initiate a next turn in turn unit 5 (A continues), or speaker B continues on talk that does not directly deal with particular issues in current talk, e.g. a new topic. For 'Ndts' there would be evidence of an uptake on current talk, for example in the form of an explicit/elaborate (dis)agreement or (dis)affiliation, and also by displaying news receipt (e.g. Heritage, 1984) or some form of appreciation. In summary, 'dts' are used as labels for cases where a hearer passes up on an uptake for a second time, whereas 'Ndts' is a label for cases where a hearer projects uptake on a second opportunity to do so. The distinction same and new topic corresponds to uptake and non-uptake (respectively) on current talk. This distinction will become clearer with the use of set of examples.

4.2.2 Illustrative examples

Four examples will be presented in this subsection, the first three illustrating cases labelled as 'doing the same', and the final one illustrates a case of 'NOT doing the same'. The phonetic features of these examples will be illustrated after having focussed on the action distinctions.

'Doing the same'. Prior to the excerpt of example 4.5 below, Lars has announced Barcelona as his favourite city in Europe, and in lines 01-05 he gives a positive assessment of a particular part of the city. In 05 Lars adds an increment to his turn at 01-02. This increment may treat Bengt's minimal response in 03 as deficient. That is, an assessment makes relevant a second assessment (Pomeranz, 1984), or an explicit appreciation of some kind, and a response of such kind is so far (03) absent. In this context Lars' increment provides a second opportunity for Bengt to respond with more affiliation. However again, rather than explicitly agreeing or elaborating on Lars' assessment, Bengt initiates a new sequence in 07 (addressing Lars' claimed access to provide the assessment).

```
(4.5) KTH-NO, BL, 17:44 "bra restauranter"

01     L:           ^DET e:r dnh° ikke spesielt   mye   som  hender ˇDE:R,
                     IT  IS       NOT ESPECIALLY MUCH  THAT HAPPENS THERE
                     there is      not a lot that happens there

02           1->    men det e:r (dnh°) (--) ↑det er veldig trivelig ↓`OMråde.
                    BUT IT IS              IT IS VERY   PLEASANT    AREA
                    but it's (--)          it's (a) very pleasant/nice area

03     B:    2->    mm.
                    mm

04                  (-)

05     L:    3->    med ´BRA  restau↓`RA[NTer] og-
                    WITH GOOD RESTAURANTS     AND
                    with nice restuarants and

06     B:    4->                        [mm. ]
                                          mm

07     B:    5->    har  du  brukt mye `TID  i: barcelona ↑↑ell`er;
                    HAVE YOU SPENT MUCH TIME IN BARCELONA   OR
                    have you spent a lot of time in Barcelona or
```

Thus, in this example there is no evidence that Bengt treats Lars' increment (05) as any different from its host (02), which is why Bengt's two *mm* responses are labelled as 'doing the same'. And by doing so, Bengt is in a sense sequentially deleting Lars' increment (turn unit 3), by attending to matters that are not directly related to its content. The 5 turn units presented above are indicated with arrows in the transcript.

Example (4.6) is also illustrative of two responses 'doing the same'. However, the shift in this example is less disjunctive than in example 4.5. Here Oscar and Anne talk about problems involved with speaking Norwegian, given that they live abroad and have partners and children who do not speak much Norwegian. Prior to, and during the excerpt below, Oscar refers to telephone conversations he has with his daughter, who lives in England and for whom English is a native language. According to Oscar he and his daughter start their conversations in Norwegian, but because his daughter's Norwegian is quite limited, they turn to English after a while.

```
(4.6) KTH-NO, AO, 17:21 "over til engelsk"


01   O:            så merker ˆJEG at eh:: nå  nå  tror  jeg ikke riktig at
                   SO NOTICE  I   THAT    NOW NOW THINK I   NOT  REALLY THAT
                   then I notice that uh now now I don't think that

02                 jeg klarer å ´SI `dette; (.)
                   I   MANAGE TO SAY THIS
                   I will be able to say this (.)

03                 °hhh på en sånn måte    at   hun forstår-=
                       ON A  SUCH FASHION THAT SHE UNDERSTANDS
                   °hhh in such a way that she'll understand

04                 =jeg underverˀ       vur¯DERer henne an[tagelig li]tt; hh°
                    I   UNDER(ESTIMATE) ESTIMATE  HER    PROBABLY   A-LITTLE
                   I probably underest- estimate her a little bit

05   A:                                                    [mm,       ]
                                                            mm

06   O:     1->    (°hh) så går vi over til `ENG´elsk,
                          SO GO  WE OVER TO    ENGLISH
                   then we switch to English

07   A:     2->    mm,=
                   mm

08   O:     3->    =etter noen mi`NUTT´er,
                    AFTER SOME MINUTES
                    after a few minutes

09   A:     4->    mm;
```

```
10                       (--)

11   A:    5->   °ptkh det  fins jo  `MANGe årsaker til å gå over til
                     THERE ARE  part MANY REASONS TO  TO GO OVER TO
                 °ptkh there are many reasons to switch to

                 `ENG´elsk,
                 ENGLISH
                 English

12               de:t er et ˇMYE ˇRIKere ´SPRÅK?
                 IT  IS A   MUCH RICHER  LANGUAGE
                 it's a much richer language
```

It appears that Anne is maintaining hearership with her second response (09), but after a gap following her second response (10), Anne continues on a tangential topic, addressing how it makes sense to choose English as a shared language (11). Although Anne's line 11 is clearly tangential to Oscar's talk and actively connects with it (notice for example Anne's lexical-grammatical repetition in *gå over til engelsk* at the end of line 11, compared to Oscar's 06), it does not elaborate on the particulars of Oscar's talk; namely how Oscar and his daughter shift from speaking Norwegian to speaking English (due to her abilities to speak Norwegian). Instead, she addresses more general issues involved. Furthermore, Anne does not show any particular orientation to Oscar's increment (08) in her next turn. As such Anne's two responses where categorised as 'doing the same'.

In a third example of 'doing the same' (4.7), Anne clearly disengages from Oscar's talk, and this is observable during her production of the second response. Here Oscar and Anne have been talking about a Norwegian female weather-forecaster on Swedish television, who according to Oscar speaks Swedish almost perfectly. This is presented as particularly news-worthy as the forecaster comes from Bergen, a city in western Norway.

Oscar's construction in 02 (1->) is a complex sentence (i.e. using *at*/"that"), however there are features in Oscar's production of 02 indicating that he does not clearly project a turn continuation. Line 02 ends with the anaphoric expression *det*/"it/that", which here does not seem to project a further turn production, since it is produced as ending the previous intonation phrase rather than initiating a new one: It is produced with a slight fall in pitch and is quieter than previous talk. Thus it seems like some display of

understanding (from Anne) is relevant at the end of 02, rather than a continuation from Oscar. I have indexed lines 06 and 07 with arrows 5a-> and 5b-> respectively, since both participants continue in turn unit 5, but (as we will see) in different ways.

```
(4.7) KTH-NO, AO, 34:22 "uforståelig"
01    O:            jeg kan ´HØRe at  de:t at  det er `NORSK    ´der,
                    I   CAN HEAR THAT IT   THAT IT  IS  NORWEGIAN THERE
                    I can hear that there is Norwegian there

02          1->     men at  hun skulle komme fra ´BERGen `det;=
                    BUT THAT SHE SHOULD COME  FROM BERGEN THAT
                    but that she should come from Bergen (that)

03    A:    2->     =<<breathy> nei,>=
                               NO
                               no

04    O:    3->     =°hh det  er `HELT   eh:: <<p> ufor`STÅelig;>
                        THAT IS  COMPLETELY      INCOMPREHENSIBLE
                     °hh that is completely uh incomprehensible

05    A:    4->     <<breathy> (ja/nei),>=
                               (YES/NO)
                               (yes/no)

06    O:    5a->    =og  `ASKøy Hole: det  lyder <<p> jo   ikke `ØSTnorsk
                     AND *name*       THAT SOUNDS     part NOT  EAST-NORWEGIAN
                    and Askøy Hole doesn't sound like East Norwegian

                    akku´rat,>=
                    EXACTLY
                    exactly

07    A:    5b->    =°h kan `DU ↑snakke ¯FLE:Re: `NORSKe   dia`lek´ter?
                        CAN  YOU SPEAK   MORE    NORWEGIAN DIALECTS
                     °h do you speak other Norwegian dialects

08    O:            nei.
                    NO
                    no
```

In 01-02 Oscar presents as incredible the fact that someone from Bergen can learn to speak perfect Swedish (*at hun skulle komme fra Bergen*/"that she should come from Bergen"). This projects an agreement, or appreciation from Anne. Anne's first response, *nei*/"no" (03) displays alignment with Oscar; however she makes no explicit treatment available at this point. Oscar then continues on his turn in 04, but Anne does little following her second response to show that she is going to do something different from before, in terms of Oscar's talk.

When Anne talks next (07) she does so on a new topic, as if there was nothing more to say about the female forecaster. This forms one kind of evidence for Anne's disengagement with Oscar's' talk. But there are two further observations that add to this claim. The first observation is regarding the production and lexical identity of Anne's second response compared to the first response. The second response (05) is transcribed as *(ja/nei)*/"(yes/no)" because it is ambiguous whether it is a *ja*/"yes" or a *nei*/"no". This is interesting because a *ja* would be fitted in terms of Oscar's positive formulation, but the fact that Anne does not unambiguously produce a *ja* suggests that she is already disengaging with Oscar's talk. The second observation is that when Oscar continues talking, in 06, he stays on the topic of the Swedish forecaster. This turn-continuation seems like a further attempt at building support for the news-worthiness of his talk (i.e. the name of the forecaster does not sound as if it is from Eastern Norway); and by placing the turn-continuation immediately following Anne's second response, Oscar treats her actions as non-projective of an uptake on his talk.

'NOT doing the same'. The examples above will be contrasted with hearer's conduct in example 4.8 below. Like example 4.5 and 4.7, the two hearer responses in example 4.8 respond to an ongoing assessment. As in those examples, an increment provides hearer with the increasing opportunity to initiate an appreciation or agreement with the speaker. In contrast to examples 4.5 and 4.7, the hearer (Lars) initiates an uptake on Sigurd's assessment, albeit resisting it.

Sigurd and Lars are talking about some of their favourite bands. Lars disagrees with Sigurd about the quality of a Swedish prog-rock band. In 01/04 Sigurd attempts to build a stronger case by reference to the band's first record. In this example, *altså det*/"you see it", was analysed as part of turn unit 1, mainly based on its intonational connection with the earlier parts of line 01 (i.e. part of the same intonation phrase).

```
(4.8) KTH-NO, SL, 13:55 "sterke saker"

01    S:    1->    den `førsteskiva  gjorde `VIRKelig inn(p)trykk på ´MEG
                   THAT FIRST-RECORD MADE    REAL      IMPRESSION ON  ME
                   that first record really made an impression on me

      1->             altså [de:           ]:t;
                      (SO)  IT
                      you see (it)

02    L:    2->    <<breathy>  [(n)ja(nh°),]>
                               YES
                               yeah

03                 (--)

04    S:    3->    det e::r  (vm) (--)`STERK´e `sak´er,
                   IT  IS               STRONG   STUFF
                   it's  (--)           great stuff

05                 (-)

06    L:    4/5->  (n)jo, <<smile> nja men det `ER vel `DET  da.>
                   YES            YES BUT IT  IS part THAT THEN
                   yes but I guess it is then ((smiling))
```

In 01 Sigurd provides an assessment of the band's first record (note that also in this example, an anaphoric expression *det*/"it/that" is used in a manner labelled as TCU-final). Lars aligns minimally in 02. Next, in 04, Sigurd adds a second part of his assessment with *det er sterke saker*/"it's great stuff". Following this, Lars initiates an explicit agreement, designed as being resistant. The resistance is displayed in the construction *men det er vel det da*/"but I guess it is then", implying that he reluctantly agrees (notice however in the transcription that Lars may soften his stance with a smile). Also, adding to this resistant agreement, Lars initiates his two responses with a nasal (i.e. *nja/njo*, instead of *ja/jo*), which make them sound like somewhat hesitant"yes" and "no" responses. Nevertheless, Lars displays explicit agreement, and shows that he treats Sigurd's talk differently on his second opportunity to respond (06), compared to the first opportunity (04). This forms the basis for labelling this instance as 'NOT doing the same'. Additional evidence for this is the fact that Lars uses Sigurd's last contribution, *det er sterke saker*/"it's strong stuff" as an explicit starting point for his agreement, *men det er vel det da*/"but I guess it is then", displaying his direct uptake on current talk.

Phonetic features. These examples suggest that there are certain phonetics associated with response tokens that are 'dts', i.e. not projecting an uptake, as different from when 'Ndts', i.e. projecting an uptake. Figure 4.A below gives phonetic representations of the response pairs in examples 4.5-4.8, showing that whereas the second response in examples 4.5-4.7, have slightly lower pitch peak and is quieter than the first response, the second response has opposite relations to the first response in example 4.8 (i.e. the second response is louder (less breathiness), and has higher pitch peak than the first response). Furthermore, the second response in example 4.7 has more central (vowels) and open (consonants) articulations than the first response, whereas the second response has more peripheral (vowels) and closed (consonants) articulations than the first response in example 4.8. Note then, that the action type labelled as 'doing the same', is not associated with 'same' phonetics, comparing first and second responses: Response pairs in both 'dts and 'Ndts' cases are associated with 'different' phonetics, but the differences seem to be ordered differently. The phonetic characteristics of 'dts' and 'Ndts' will be addressed further in the next section.

Example 4.5	Example 4.6
[m] [m̥]	[m̥] [m̥]

Example 4.7	Example 4.8
[n̥ɛɪ] [ʒ̃ẽ̃]	[ŋ̥əä] [nju]
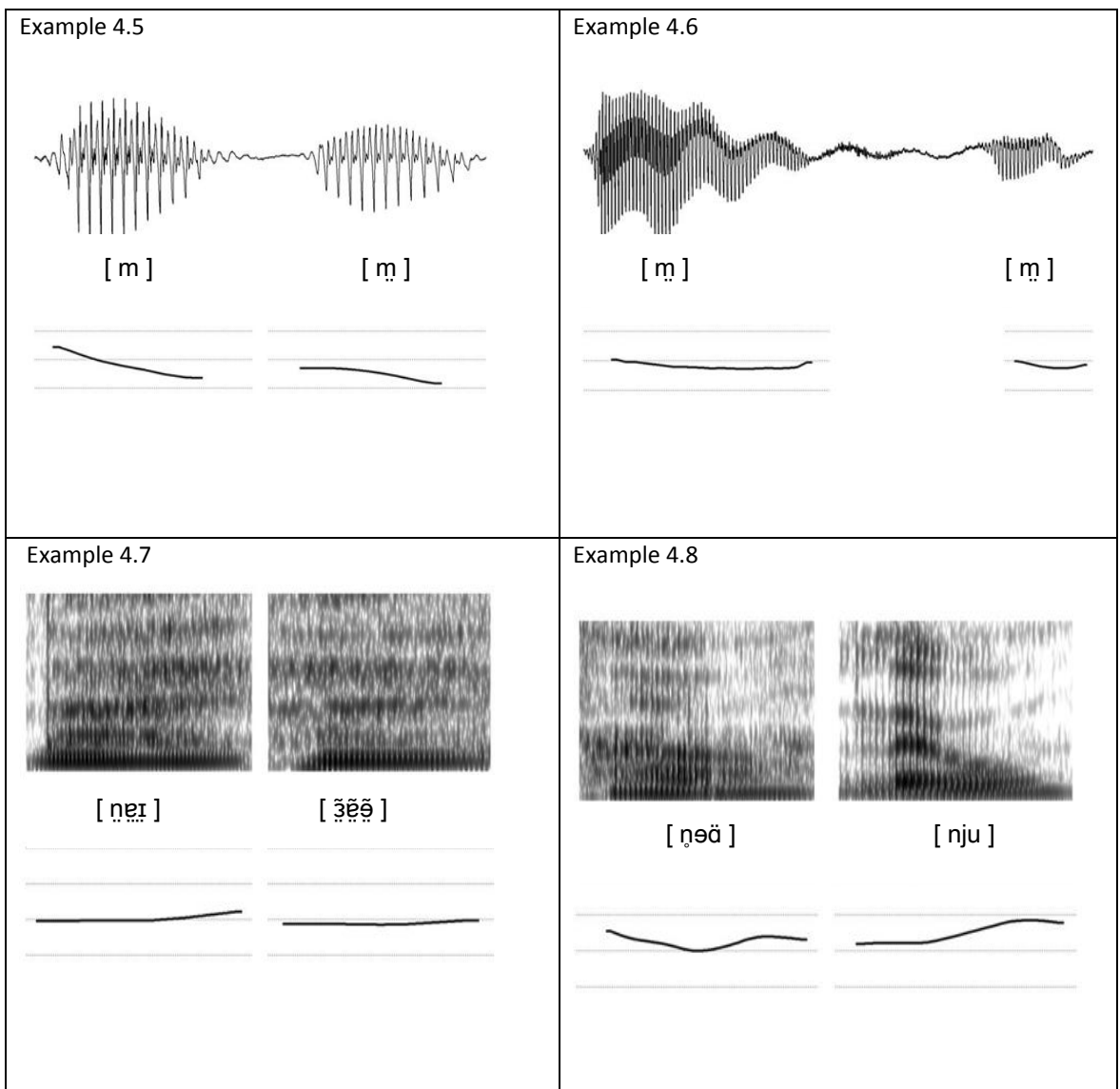
Figure 4.A. Phonetic representations of response pairs. Spectrogram (0-5000Hz), IPA transcription and pitch trace (in semitones (st); distance between each horizontal line is 5 st), of response pairs in examples 4.5-4.8.

4.2.3 Distribution of action categories according to next turn and lexical tokens

Turn unit 5. Overall there were 49 sequences found to match the criteria described in Table 4.A. 28 of these instances were labelled as 'dts', whereas 21 were labelled 'Ndts'. The distributions of these categories are presented here according to what happens following the second response (i.e. in turn unit 5).

The distribution of 'dts' and 'Ndts' shows that neither category is deterministic in terms of who speaks next, and whether it is on topic, but nevertheless suggests that 'Ndts' is associated with speaker change, on topic, whereas 'dts' is associated with either speaker A continuation (on topic), or speaker B topic shift (see Figure 4.B).
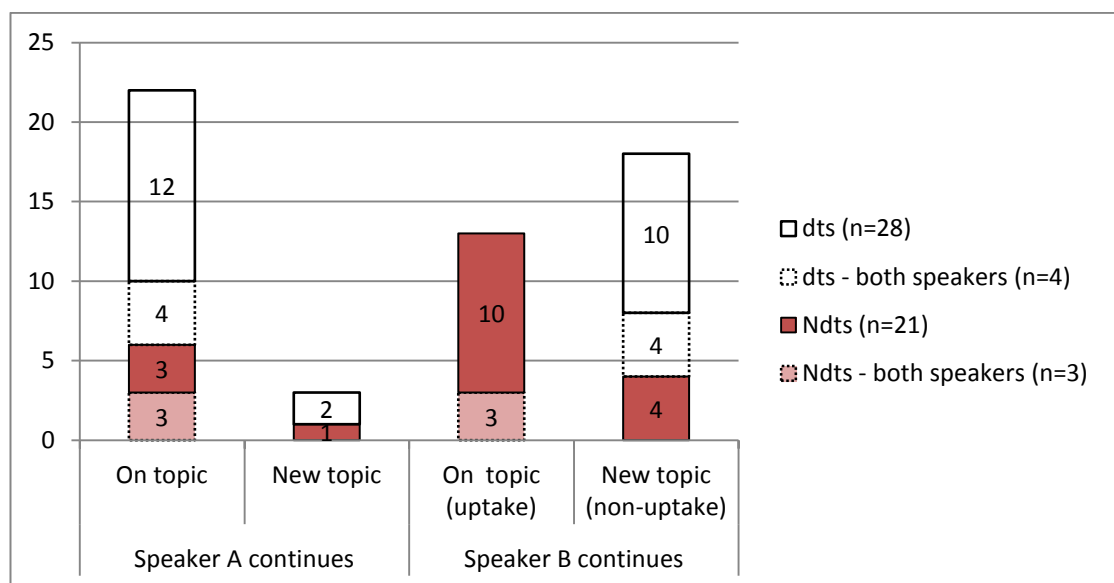
Figure 4.B. Number of instances labelled 'dts' (transparent) and 'Ndts'(coloured), according to (i) who speaks in turn unit 5, and (ii) whether or not the speaker continues on current topic. In a total of seven instances turn unit 5 is produced by both speakers (explaining the total of 56 instances in the columns, i.e. n=49+7). These are indicated with dotted lines.

Corresponding with the pre-defined criteria above (section 4.2.1, pp. 73-76), all instances of 'dts' where speaker B produces turn unit 5, were on a different topic (n=14). These include four cases where both A and B continue (marked with dotted

lines). Unlike B, A typically stays on topic, both in 'dts' and 'Ndts' cases. This suggests that 'dts' is a resource for hearer to disengage with current talk, but only if the main speaker does not find it relevant to continue on topic.

'Ndts' is mainly associated with a speaker change on topic. However, in a total of seven cases, speaker A continued following 'Ndts' (including three cases where both speaker A and B continued on the same topic, marked with light red), and in 4/17 'Ndts' cases speaker B continued on a different topic. These instances were commonly brief appreciations, which by speaker B were followed by a tangential topic, i.e. not the kind of disjunctive topic shifts found among the 'dts' cases.

Response tokens. The distribution of response tokens suggests that there is some pattern to what lexical type of responses occur when 'dts' (typically *mm-mm*) compared to 'Ndts' (typically *ja-ja*/"yeah-yeah"). However, both lexically similar and dissimilar sequences of response tokens are found within the two action categories. The distribution is given in Table 4.B. Notice that there are no cases of *ja* followed by *mm* in the collection. If Gardner (2001) is right in viewing "mm" as a 'weaker' and more disengaging response than oral responses (e.g. "yeah"), and if we assume that this is the case also in Norwegian, one explanation for the lack of *ja – mm* response pairs might be that a hearer is already heading towards disengagement when producing the first response in most 'dts' cases. But this does not explain the occurrence of *ja – ja* pairs among the 'dts' cases, and there is indeed a case of *nei* followed by *mm*. Thus one might conclude that 'dts' is not clearly defined by particular lexical responses, or whether or not the second response token is the same lexical item as the first response.

Table 4.B. Inventory of response pairs in collection, for interactional categories 'dts' and 'Ndts'. Translations: *ja* - "yes", *nei* - "no", while *mm* resembles the English "mhm" and "mm", and *øhø* resembles the English "uhuh" (in phonetic terms).

Response pair	'Doing the same'	'NOT doing the same'
mm - mm	**10**	-
mm - ja	4	3
øhø - ja	2	-
nei- mm	1	-
ja – ja	3	**11**
nei - nei	1	1
ja - nei	2	1
nei - ja	3	1
other (e.g. 'okay')	3	4
TOTAL	**28**	**21**

4.2.4 Summary

This section has presented the interactional basis for separating hearer response action according to whether or not they are 'doing the same'. Instances of a particular sequence type were collected, where a hearer responds twice to consecutive elements (host + increment) of a turn. The interactional analysis addressed whether or not hearers, in terms of action, maintained or differentiated their responses to those consecutive elements, described as 'doing the same', and 'NOT doing the same'. The interactional evidence for the distinction between these categories was provided on the basis of case-by-case analysis (4.2.2). This analysis showed how the distinction relates to hearer's displayed uptake ('NOT doing the same') or non-uptake ('doing the same') on current talk. The distributional data (4.2.3) gave supporting evidence of the interactional relevance of 'doing the same' compared to 'NOT doing the same', showing that the two are associated with different interactional consequences: Speaker A typically continues after 'doing the same', and speaker B typically continues after 'NOT doing the same'. When speaker B continues after 'doing the same', they do so on a

different topic, also when speaker A in overlap continues on the same topic. This shows that when a hearer makes an undifferentiated action, this may or may not result in a speaker change, but when it does the next turn is on a new topic. Further evidence, i.e. for how the distinction drawn above is real to the interactants, will be provided in section 4.4. Here I will turn to further interactional analysis involving deviant case studies.

There were some tendencies of particular lexical pairs being associated with particular actions, however both similar and dissimilar response pairs were used in both action categories.

4.3 Phonetic analysis

The response pairs in each sequence collected were labelled according to whether they qualified as phonetically 'less' (i.e. not 'more') or 'more' (i.e. not 'less'), based on initial case-by-case observations.

The hypothesis was that:

- Phonetically 'less' responses correspond to 'doing the same', whereas phonetically 'more' responses correspond to 'NOT doing the same'

The background for the phonetic definitions of 'less' and 'more' will be described in section 4.3.1, and the findings will be presented in section 4.3.2.

4.3.1 Procedures

The phonetic parameters analysed were:

- Voice quality
- Duration
- Pitch: mean, range, movement

- Articulation

- Loudness

These parameters were meant to cover a range of phonetic detail that might be of relevance to different extents, or in combination. Attending to these parameters, a pair-wise comparison was made between the first and second response, for each sequence separately. Each response pair was then labelled according to a phonetic comparison, in which the phonetic labels 'less' and 'more' were given complementary definitions. The motivation behind these complementary labels was based on initial observations, where it appeared that a response pair that in action terms were 'doing the same', had certain phonetic features in common as different from those used when 'NOT doing the same'. In terms of pitch for example, a second response seemed to have similar or lower pitch mean when 'doing the same', and substantially higher pitch mean when 'NOT doing the same'. Also, a second response seemed to be quieter when 'doing the same', and louder when 'NOT doing the same'. Such observations formed the motivation for naming the labels 'less' (i.e. not 'more'), and 'more' (i.e. not 'less').

Difference in loudness and pitch was decided on the basis of 'just noticeable difference' (e.g. Moore, 1989), whereas difference in voice quality and articulation was based on impressionistic listening. The criteria are summarised in Table 4.C below. Further descriptions of the phonetic analyses are provided below, ordered by phonetic parameter.

Statistical analyses were not performed in this study, as a majority of phonetic comparisons were made impressionistically, and because the analysis was restricted to a small data set.

Table 4.C. Complementary criteria for labelling response pairs as phonetically 'less' or 'more'. The definitions compare the second response with the first response, for each phonetic parameter separately.

PHONETIC PARAMETER	EXPECTED PHONETIC CHARACTERISTICS OF THE SECOND RESPONSE COMPARED TO THE FIRST RESPONSE	
	Phonetically 'less'	Phonetically 'more'
Voice quality	Similar, or less modal	More modal
Duration	Same, or shorter	Longer
Pitch mean	Same, or lower	Higher
Pitch range	Same, or narrower	Wider
Pitch movement	Same	Different
Articulation	Similar or more central (vowel) and open (consonant) articulation	More peripheral (vowel) and closed (consonants) articulation
Loudness	Same, or quieter	Louder

The analysis was done by combining acoustic analysis using software Praat (see chapter 3) and impressionistic listening. The analyses of voice quality, pitch movement, articulation and loudness were primarily based on careful impressionistic listening, whereas duration, pitch mean and pitch range was based on instrumental analysis.

Voice quality. Voice quality was assessed in terms of phonation, which in physiological terms is related to the degree of tension across the vocal folds and their mode of vibration (e.g. Laver, 1994). Modal phonation, with moderate tension (as in 'normal' speech), was distinguished from creaky (more tension) and breathy (less tension) phonation. For example, if both responses in a response pair were produced with modal voice quality, the case was labelled as phonetically 'less' (i.e. 'not more'). If the second response had more breathy or creaky voice quality than the first one, it was also labelled as 'less'. But if the first response had more breathiness/creak than the second one, the case would be labelled as 'more' (i.e. the second response is more modal). In the pair-wise comparison breathy and creaky phonation types were regarded as equal in terms of 'less modalness'.

Duration. Duration was measured manually in Praat. The boundaries were set to include only the audible portions of the response, and excluding following outbreath etc. The

pair-wise comparison was done on the basis of absolute (linear) values, i.e. judgments on whether two responses were 'not more', shorter or longer, represents same, shorter or longer values in absolute terms.

Pitch mean and range. Pitch mean and pitch range were measured manually in Praat. Within the boundary set for duration, only the portions that were produced with a stable phonation were included. This normally excluded the first and last couple of periods in the waveforms. Each trace was investigated and corrected according to Praat-errors in judging voicing, and perturbations due to creaky voice. Overall mean values were determined on the basis of the resulting pitch trace, and range values were based on the highest and lowest peak points in the pitch trace. Mean and range value were converted into semitones, because semitone is more closely related to perception than absolute Hz values are (Nolan, 2003). The pair-wise comparisons were made with reference to the speaker's overall pitch range, decided on the basis of the highest and lowest pitch points during 5 minutes for each speaker, collected from equal intervals in the recordings. If the semitone values were close to identical, judgments on 'less'/'not more' or 'more' were based on impressionistic listening to any noticeable difference.

Pitch movement. Pitch movement was based mainly on impressionistic cues as to the direction of pitch movement during the response token. The main categories were fall, rise, fall-rise and rise-fall.

Articulation. Assessment of articulation was based on impressionistic listening. Only those responses produced with an oral airstream were assessed in terms of articulation, excluding "mm" responses from the analysis for this parameter. For vowels, judgments pair-wise comparisons were based on the closeness to cardinal vowels, i.e. how peripheral the vowel quality was, in view of what a canonical version of that response token might be. For consonants, judgments and comparisons were based on the closeness of constriction.

Loudness. Loudness judgments were based mainly on impressionistic listening, and pair-wise comparisons were based on whether or not there was any noticeable difference between the two consecutive responses.

4.3.2 Findings

The number of matches between 'dts' sequences and response pairs labelled phonetically 'less' varied across the phonetic parameters, ranged from 57-100% (see Figure 4.C). The highest match scores were found for voice quality (100%), loudness (92.9%), and articulation (92.3%), making these seem the most reliable phonetic cues to 'dts'. The lowest match score was found for pitch mean (57.1%) and articulation (60%).

The number of matches between 'Ndts' sequences and response pairs labelled phonetically 'more' showed an overall lower match score than between 'dts' and 'less' phonetics, ranging from 33.3% (duration) to 81.0% (vowel quality).
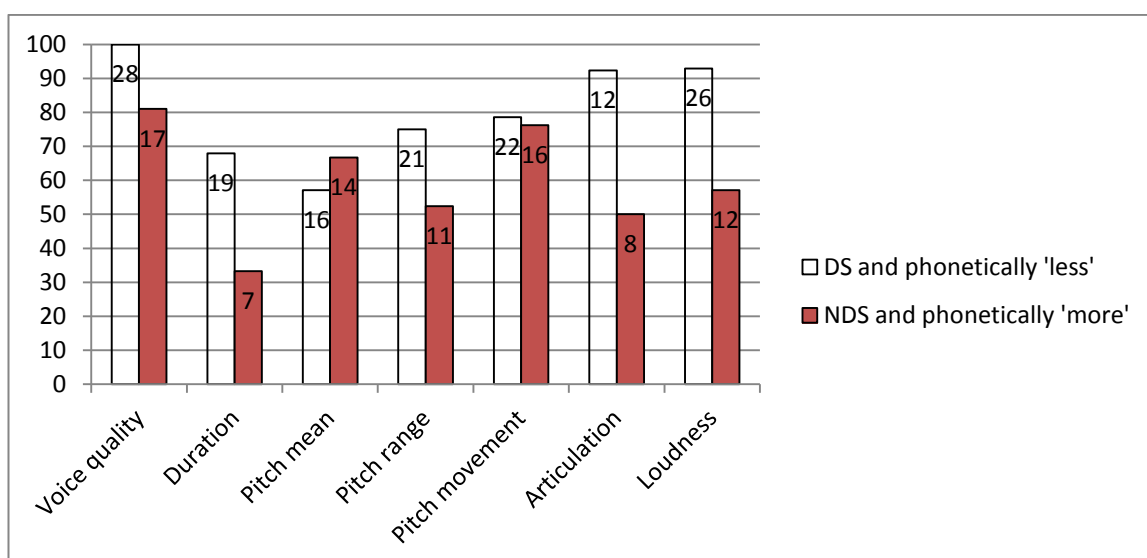
Figure 4.C. Correspondence (in percentage) between response pairs labelled phonetically 'less' with 'dts', and response labelled phonetically 'more' with 'Ndts', for each phonetic parameter. N instances are given for each column. Total N=28 for 'dts' and 21 for 'Ndts', for all phonetic parameters except articulation, where N=13 in 'dts' and N=16 in 'Ndts' cases (due to *mm* productions).

The findings presented in Figure 4.C show that there is some regularity to the phonetic characteristics in 'dts' sequences, but that these characteristics overlap to some extent with those for 'Ndts', within individual parameters. However, post-hoc analyses

revealed some further complementarities between the categories, in that the non-matching phonetic features typically followed a regular pattern. These further complementarities and overlaps between categories are summarised in Table 4.D.

A majority (9/12) of non-matches between expected and observed pitch mean for 'dts' cases (i.e. cases where the second response had higher pitch mean than first response) for example, were found to be within 9% of the speaker's overall pitch range. On the other hand, 10/14 *matches* for 'Ndts' (i.e. cases with the highest pitch mean in the second response) were found to be above 9% of the speaker's overall pitch range. Also, in the majority (6/7) of non-matches between expected and observed pitch means for 'Ndts' (i.e. cases where the second response had the same or lower pitch mean than first response), the second response had more than 10% lower pitch mean, with reference to the speaker's overall pitch range. Only 1/16 of 'dts' cases had such a substantially lower pitch mean in the second response (i.e. 15/16 cases of lower pitch mean was found within 10%). In other words, there is some basis for including slightly higher pitch means in the 'less' phonetic category, whereas substantially lower pitch means may signal 'more' phonetics.

Furthermore, focussing on what is found typically *not to* occur phonetically in 'dts' and 'Ndts' cases, rather than on strict complementarities, also reveals some regular patterns. For example, it was *not found* in 'Ndts' cases that a second response would have more open consonantal articulation and more central vowel articulation than a first response, and the reverse was true for 'dts' (with one exception). Similarly, in 'dts' cases only 2/28 cases had a noticeably louder second response than a first response, whereas in 'Ndts' cases only 2/21 cases had a noticeably quieter second response than a first response.

Table 4.D. Summary of observed phonetic characteristics for 'dts' and 'Ndts', based on post-hoc analyses. Complementary features are given in black, with numbers and percentage observed cases. Features overlapping between the two action categories, and observed numbers/percentages, are given in grey. Phonetic parameters are indicated in the left-most column: Voice quality (VC), duration (Dur), pitch mean (PMe), pitch range (PR), pitch movement (PMo), articulation (Art) and loudness (L).

OBSERVED PHONETIC CHARACTERISTICS OF SECOND RESPONSE COMPARED TO FIRST RESPONSE

	'dts'	N	'Ndts'	N
VC	< modal, or	28/28 (100%)	> modal, or	20/21 (95.2%)
	similar (/total)	3/28 (10.7%)	similar (/total)	3/21 (14.3%)
Dur	Shorter, or < 25% longer	25/28 (89.3%)	Longer, or > 25% shorter	15/21 (71.4%)
	> 25% shorter (/shorter)	2/19 (10.5%)	< 25% longer (/longer)	0/7 (0%)
PMe	Lower, or < 9% higher	25/28 (89.3%)	Higher, or > 10% lower	20/21 (95.2%)
	> 10 % lower (/lower)	1/16 (6.3%)	< 9% higher (/higher)	4/14 (28.6%)
PR	Narrower, or < 3% wider	24/28 (85.7%)	Wider, or > 20% narrower,	16/21 (76.2%)
	> 20% narrower (/narrower)	3/21 (14.3%)	< 3% wider (/wider)	1/11 (9.1%)
PMo	Similar, or reduced version	24/28 (85.7%)	Any different (but not reduced version)	18/21 (85.7%)
		(0%)		(0%)
Art	< closed (consonants)/ < peripheral (vowels), or	12/13 (92.3%)	> closed (consonants)/ > peripheral (vowels), or	16/16 (100%)
	no noticeable difference (/total)	3/13 (10.7%)	no noticeable difference (/total)	8/16 (50%)
L	Quieter, or	26/28 (92.9%)	Louder, or	19/21 (90.5%)
	no noticeable difference (/total)	11/28 (39.3%)	no noticeable difference (/total)	7/21 (33.3%)

This analysis focuses on single phonetic parameters rather than on a combination of them. A second part of the post-hoc phonetic analysis reveals that the amount of non-matches between phonetic categories ('less' and 'more') and action categories ('dts' and 'Ndts') is limited to only a few phonetic parameters for any single case. These are summarised in Table 4.E. In any single case the non-match is found in half or less of the total amount of phonetic parameters. Most of these cases have non-matching features in one phonetic parameter only (18/28 cases), while some cases show non-matching features for two (7/28 cases) and three (3/28 cases) phonetic parameters, but in no case with more than three, out of the total of seven parameters.

This may support an argument for focussing on a bundle of phonetic features, in a case-by-case manner, rather than separate phonetic features (e.g. Local, 2004). That is, there is no strong indication that only one or a few phonetic parameters are associated with the action distinction reported on here, and that the others are irrelevant. To further address this issue one might attempt to tease apart different parameters in an experiment, which might also be used to test whether or not there is a certain perceptual threshold between 'less' and 'more' phonetic features (see also discussion, section 4.5).

Table 4.E. Overview of non-matches between phonetic categories. The overview gives the total number of cases (second column), and the amount of deviant phonetic features involved (columns 3-6).

Action category	N cases with non-matching phonetic features	N non-matching phonetic features in any single case (7 phonetic features in total)			
		1	2	3	> 3
'dts'	15 (of 28)	10	4	1	0
'Ndts'	13 (of 21)	8	3	2	0

4.3.3 Summary

The phonetic comparisons between response tokens that are 'doing the same', and response tokens that are 'NOT doing the same', reveal that there are some regularities

distinguishing the two action categories. When 'doing the same', the second response typically has/is:

- Similar or more breathiness/creakiness in terms of voice quality (*never* more modal)
- Shorter, or <25% longer in terms of duration
- Lower pitch mean (or <9% higher)
- Narrower pitch range (or <3% wider)
- Similar pitch movement
- More central/open vowel/consonant articulation
- As loud or quieter

than the first response. When 'NOT doing the same', the second response typically has/is:

- More modalness in terms of voice quality
- Longer, or >25% shorter in terms of duration
- Higher pitch mean (or >10% lower)
- Wider pitch range (or >20% narrower)
- Different pitch movement
- More peripheral/closed vowel/consonant articulation (*never* less peripheral/closed)
- As loud or louder

than the first response.

There are overlaps within each phonetic parameter. But rather than looking at strict complementarities for phonetic parameters separately, it might be more useful to think about these results in terms of a combination of features. It was shown that 'dts' and 'Ndts' always match more than half of the phonetic features expected to correspond with these categories (i.e. there were no less than 3 mismatches). Thus, it might be that the absence of one phonetic feature might be compensated by the presence of others.

4.4 Further interactional analysis

Having found that certain phonetic characteristics typically go along with response tokens that are 'doing the same' in terms of action, this can be used to conduct further interactional analysis. This section presents two instances that for different reasons did not fit with the sequence explored above, but which nevertheless are informative of the practice at hand.

4.4.1 'Doing the same' in longer than minimal units

Example 4.9 has some similar features to those presented above, in that a hearer responds twice, before and after speaker's increment to his turn. But we will see that phonetic resources for 'dts' can also work over longer units than a minimal response. Also, this is oriented to more explicitly as lack of engagement with the current talk, compared to the examples above.

Lars and Tor are acquaintances who grew up in the same town in Norway but didn't know each other at that time. Prior to the example Lars has picked up on information about the time Tor left town to start his University degree. Lars requests Tor to specify when he graduated from high-school in Norway, which was in 1991 (line 1). In 01 Lars reveals that he himself graduated in 1992. The main issue here for the interactants is whether they both went to the same high-school at the same time; an understanding they eventually make explicit in 07-08. Of main analytic interest is how Tor does not seem to engage with such an understanding at first, but then if 'forced to' by Lars.

```
(4.9) KTH-NO, TL, 05:44 "fra Larvik"

01    L              [j ]eg gikk ut  våren? (.)ˀn hˈ våren-
                     I     WENT OUT SPRINGdet       SPRINGdet
                     I graduated the spring uh the spring

                     ´ja `våren heh heh ˈhh nitti:´T:O: `tenker jeg;
                     YES SPRINGdet        NINETY-TWO   THINK  I
                     yes the spring ((laughter)) ninety two I think

02                   (--)

03    T              ja ´Ok`ay.
                     YES OKAY
                     yes okay

04    L              fra  larvik;
                     FROM name
                     from Larvik

05                   (--)

06    T              ja ´DER   `ser man;
                     YES THERE  SEE ONE
                     yes there you go

07    L              ((head-move)) så[: eh: heh  heh   heh   heh heh   heh   ]
                                      SO
                     ((head-move)) so: uh: ((laughter))

08    T                                [<<ff> ´DA   har  vi gått på gym`NASet heh]
                                              THEN HAVE WE GONE ON HIGH-SCHOOLdet
                                              then we went to high school huh

                     ´SAMtidig.>
                     SAME-TIMEadv
                     at the same time
```

Tor responds minimally in 03, following Lars' assertion on when he left high-school. At this point Tor does not project any uptake as to whether they went there at the same time. This is evidently an issue for Lars as he increments with *fra Larvik*/"from Larvik" (04). With this he seeks to establish that they indeed went to the same school. Again, Tor does not project any uptake (05-06), but adds the rather idiomatic and not particularly engaged *ja der ser man*/"yes there you go". Apparently then, Tor does not make the same connection as Lars, or resists displaying it at this stage.

Tor produces the second response with phonetics characteristic of 'dts': The second response *ja der ser man*/"yes there you go" (line 11) is produced with a slightly softer production, less peripheral articulation and more breathiness in the second response than in the first. And in terms of pitch, the pitch movement remains quite the same, whereas the pitch range is narrower and the pitch mean is lower in the second response compared to the first (see figure 4.D).
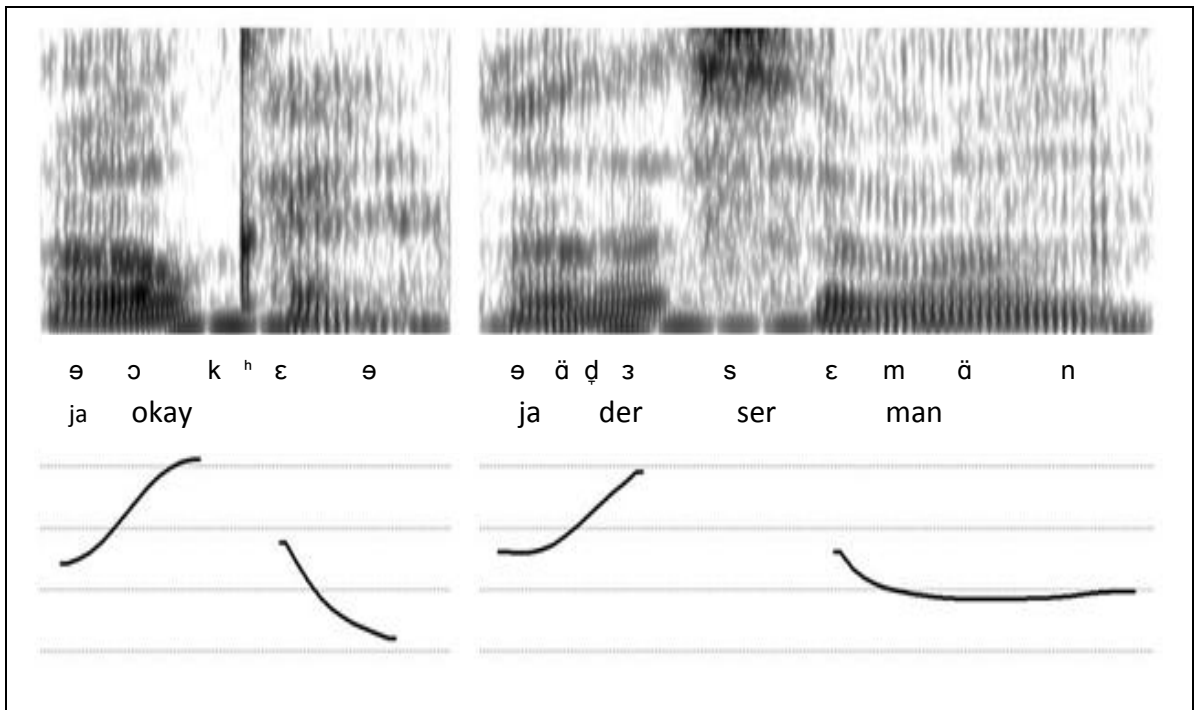
Figure 4.D. Phonetic representation of example 4.9. Spectrogram, IPA transcription and pitch contour (each line represents 5 semitones) for the two responses *ja okay*/"yes okay" and *ja der ser man*/"there you go", from lines 7 and 11 in example 4.9.

Tor's second response is out of place in terms of what it is that Lars projects with his talk. This is evidenced in how Lars next makes it clear to Tor that something more is wanted from him. Lars does this with the combination of a head-move (to the side, still maintaining mutual gaze with Tor) and *så...*/"so...". This is immediately followed by Tor's explicit formulation *da har vi gått på gymnaset samtidig*/"then we went to high school at the same time". With this production Tor seems very willing to show that he understands what Lars wants: He initiates the turn considerably louder than his previous talk (and much louder than Lars' *så*). In this way Tor avoids being later than he already is with his display of understanding.

In summary, this example shows that phonetic resources for 'doing the same' can work as a display of disengagement also in longer structures than response tokens.

4.4.2 A deviant case

The final example (4.10) seems at first to argue against the analysis, which was based on the examples in section 4.2. Here the hearer (Oscar) follows his second response with an assessment, *det er fantastisk*/"that's fantastic", but the second response clearly corresponds to the phonetic pattern common for 'dts' rather than 'Ndts'. However, as will be shown with detailed interactional analysis, this example provides further evidence to the general claim about 'dts' and 'less' phonetics. The apparent uptake is a display of disengagement rather than a genuine assessment, as evidenced by the interactants' orientations (which include some visual detail). This description will also address a second deviant feature, namely the fact that Oscar's first response occurs in the middle of Anne's turn construction (line 10). The example was not included in the core collection based on this second deviant feature.

Anne and Oscar have lived in Sweden with their non-Norwegian partners/spouses for more than two decades. Oscar has expressed that he uses Norwegian very little in his daily life, and finds it difficult to speak Norwegian with Swedes. Anne on the other hand implies that speaking Norwegian with Swedes is unproblematic for her. Prior to this sequence Anne has announced that she speaks Norwegian at home (Göran is Anne's Swedish partner), but in 01 Oscar challenges this claim.

```
(4.10) AO, 10:30 "norsk hjemme"

01    O:            ‾TIL og  med  nå   med `GÖRan  altså,
                     TO  AND WITH NOW  WITH name   THUS
                    even now with Göran (you say)

02                  (--)

03    A:            ja:,
                    yes

04                  [selv`F]ØLGe´lig,
                    CERTAINLY
                    of course

05    O:            [ja,   ]
                    yes

06    O:            ja,
                    yes
```

```
07    A:              spesi´ELT   med  `GÖran.
                      ESPECIALLY  WITH  name
                      especially with  Göran

08    O:              mm,
                      mm

09                    (.)

10    A:      ->      han snakker `BÅde: (d) [(.) ] norsk:    og ´DANSK `han;
                      HE  SPEAKS   BOTH           NORWEGIAN AND DANISH HE
                      he speaks both (d) (.)          Norwegian and Danish

11    O:      ->                                  [ja, ]
                                                   yes

12    O:      ->      ja, <<p> [det  er fantas[tisk-]>
                      YES       THAT IS FANTASTIC
                      yes that's fantastic

13    A:                       °h[h            [heh  ]
                               °hh               ((laughter))

14    A:              °hh m:en: eh: (hh) ↑jo    men altså jeg ((...))
                          BUT              part BUT THUS  I
                      °hh but uh          yeah but you see I
```

After having challenged Anne's claim to speak Norwegian at home in 01 (Anne lives with Göran, her Swedish partner), Anne answers by contesting Oscar's question, the most central element of this contestation being the use of *selvfølgelig*/"of course" (cf. Stivers, in press). Although Oscar accepts Anne's answer minimally in lines 05, 06 and 08, Anne proceeds to account for speaking Norwegian at home with Göran in 10: Apparently he speaks both Norwegian and Danish.

There are several indications that Oscar works actively to bring Anne's actions to an end (e.g. by disengaging with her talk). This could perhaps best be described as a 'counter-engagement'. First, at line 11 Oscar produces an early, minimal response to Anne's ongoing turn construction (10). By doing so, Oscar displays (i) that he knows, i.e. anticipates, what comes up next (i.e. Göran's accommodating linguistic abilities, which in terms of action further contests Oscar's question), and (ii) that he accepts Anne's claim to speak Norwegian at home. As further (visual) evidence of Oscar's disengagement, Oscar closes his eyes right before his first response, and keeps them closed until the end of Anne's turn in 10 (see for example Kendon, 1977, on how gaze is systematically associated with displayed hearership). Also, Oscar produces a slight

nodding gesture throughout Anne's turn, which might further add to the impression that Oscar want to show that he knows where Anne's turn is heading.

This argument might at first seem inconsistent with Oscar's subsequent action in 12, where he goes on to assess Göran's language abilities. However, *det er fantastisk*/"that's fantastic", in 12 is not really an assessment. It is in fact designed, and understood (by Anne), as providing a sarcastic stance towards Anne's talk. A central design feature of this is Oscar's lexical choice: *Fantastisk*/"fantastic" is clearly a strong word when taking context into consideration. That is, Norwegian, Swedish and Danish are all mutually comprehensible languages, and that speakers of these languages are able to accommodate each other is not 'fantastic'. Anne picks up on this with a laughter token in 13 to which Oscar responds with a smile (not included in the transcript). Another relevant design feature in Oscar's 'assessment' in 12 seems to be its very quiet production, involving a breathy-creaky voice quality (during which he also closes his eyes).

On this basis, it is argued that Oscar is indeed 'doing the same' with his two response tokens. As shown in Figure 4, Oscar uses phonetic characteristics that typically go along with 'dts', including quieter, more open/central articulation and a lower pitch mean in the second response, and the two responses have similar pitch movement.

[ṇja]   [ə̥əḁ̈]
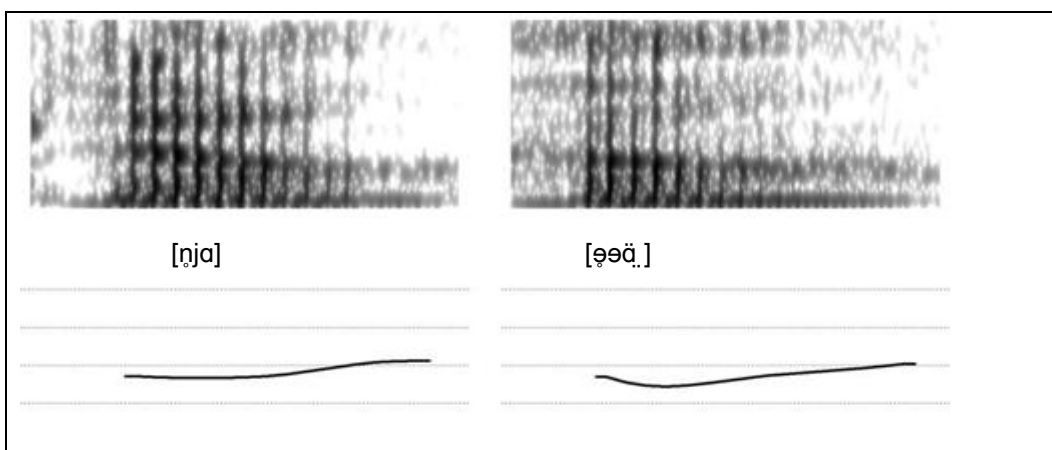
Figure 4.E. Phonetic representation of response pair in example 4.10. Spectrogram (0-5000Hz), IPA transcription and pitch trace (in semitones, the distance between each horizontal line is 5 st).

In summary, example 4.10 demonstrates some of the further activities 'dts' can take part in. The implications of 'dts', as displayed by the interactants themselves, *confirm* the general claim rather than disconfirming it. That is, rather than changing topic, Oscar accounts for his (early, perhaps inappropriate) disengagement by providing a sarcastic stance towards Anne's talk, and based on Anne's and Oscar's mutual orientations, this is seemingly done to actively 'rush' Anne's action towards an early closure, and towards the relevance of a shift in talk.

4.5 Summary and discussion

This study shows that hearers relevantly distinguish between maintaining and differentiating their actions in certain sequential circumstances, and that there are phonetic characteristics that typically go along with such a distinction. Furthermore, the study shows that hearers may actively use these phonetic resources for maintaining their actions (i.e. 'doing the same') in working towards a sequence closure, and a topic-shift.

There are important implications of this work regarding how hearers take part in disambiguating what happens next in talk in interaction; and in shaping coherence between turns, and sequences of turns. It should be noted that the practice 'dts' is not deterministic in terms of what happens next, i.e. 'dts' does not *mean* that a topic change is coming up. However, 'dts' can be used as a resource for disengagement with current talk, and as such this study adds to the literature on how topic changes are negotiated and achieved.

According to previous studies on topic organisation in talk-in-interaction (e.g. Jefferson, 1984; Jefferson, 1993; Holt and Drew, 2005), most topic transitions are organised in a seamless fashion, "without any overt termination of one prior to the introduction of a next" (Holt and Drew, 2005, p. 41). However, there are sometimes particular 'pre-shift tokens' (Jefferson, 1993) involved in shifting a topic, such as figurative expressions (Holt and Drew, 2005), items like "so" (Raymond, 2004) and the use of lexical repetition (Curl

et al., 2006). All these resources show attention to previous talk, while heading for something else, without making this shift explicit or definite. The resource for 'doing the same' contributes to these findings. 'Doing the same' might be even less explicit than the other pre-shift tokens reported on; i.e. it does not draw attention to the topic-shift as a shift and thereby provides a more seamless shift.

This study also has important implications for the study of phonetic variation/variability in spontaneous talk, and in understanding the function/meaning of response tokens. Instead of studying response tokens as single items, with isolated meanings, this study offers a distinctly different approach, putting interactional detail and development at the centre of the analysis. Since response tokens are highly contextualised items, such analysis offers a significant contribution to the ways in which response tokens take part in the interactional process, and shape meaning. Response tokens are observed in locations where e.g. agreement, affiliation, and sequence closing made relevant, and thus their design may be crucial in displaying to what extent the hearer is taking up on these relevancies. This is in line with Heritage (1984, p. 335): "these objects [response tokens] are used to achieve a systematically differentiated range of objectives which, in turn, are specifically consequential for the onward development of the sequences in which they are employed".

This chapter contributes to the understanding of *how* such systematic achievements are made, informed by careful analyses of phonetic *and* interactional details. Seemingly, phonetic characteristics are important to make the relevant distinctions in these locations, perhaps more distinctly so than lexical distinctions. A potential explanation for how phonetics are so important to distinguish response tokens, is that such a brief vocal production little room for verbal conduct, and lexical variation, and therefore their phonetic features are important to distinguish what kind of response is produced. In a full sentential turn in comparison, there is a whole proposition to do the work.

There are some implications here, for how to think about what is lexical, what forms lexical distinctions, and also what to include in a language description, e.g. a lexicon. First, I would argue that response tokens like "mm" can be categorised as lexical tokens just like "yes" or "no", in that they form certain functions in a particular interactional

context. Maintaining that these are all lexical items, it is still interesting to consider what distinguishes them as lexical items. At the same time as a "yes" and an "mm" have different phonetic-segmental features, this study shows that (non-lexical) phonetic features can be used to shape such on the surface 'different' lexical items to do the same action. And similarly, identical lexical items can be shaped to do different actions. If these items were to have an entry in a dictionary, it seems relevant to include such information, i.e. that the distinction between these lexical items (in terms of function) is influenced by non-lexical phonetic characteristics, in certain interactional environments. Again, the point is that function does not arise from on-the-surface lexical categories alone. It is possible that some of these non-lexical features are found across languages (e.g. not just in Norwegian), however they may still be relevant in a linguistic description (see also chapter 7, section 7.3).

Studying phonetics in terms of sequence and action helps accounting for more of the variability in response tokens, not previously reported. This chapter demonstrates how this kind of work can be done, by paying careful attention to both phonetic detail and to sequential structures in which such detail is systematically used. Interactional analyses led to the discovery of a type of structure (sequence), in which phonetic detail systematically makes a difference to meaning-making. This understanding was crucial for collecting a set of comparable cases, and thereby achieving analytic 'control' (and linguistic comparability). All this was done while keeping the interactants' own displayed orientations and understandings at the core of the analysis.

In future work it might be possible, and desirable, to further test the role of phonetic design in response tokens using an experimental setting. For example, one may use a controlled set of responses (automatically generated in an approximated naturalistic interactional setting), and test their consequences on the participant's next actions. In such a setting one might also manage to manipulate different sets of phonetic parameters, to test their relative importance in making interactional contrast. A good example of a study that tests the relative importance of different linguistic features in turn-taking is that of Ruiter, et al. (2006).

The chapter adds to the observations made by Jefferson (1985, 1993) and Gardner (2001), although their studies investigated response tokens in English. Jefferson (1993) suggested that different forms of "yeah" may project a speaker change, and indeed a topic-shift. This study offered a more precise description of such different forms, showing that responses associated with topic-shift ('dts') and responses associated with continuation on topic ('Ndts') may both be phonetically different from previous responses, but in different ways. The current study also indicated that difference/similarity in terms of lexical category alone is not clearly associated with particular interactional consequences. This finding is somewhat in conflict with those by Jefferson (1985) and Drummond and Hopper (1993), regarding "mm" – "yeah" sequences.

Gardner (2001) argues that (English) "mm"s are weak and somewhat disengaging response tokens, when accompanied by a falling pitch contour. As shown in the current study, a variety of lexical items may work as disengaging, and in order for a response to be disengaging, a relevant feature of disengagement is that the token is shaped similarly to a previous response. It is important to note that the current study explores a particular kind of context separate from that of Gardner (2001), and that Gardner (2001) may also have identified lexical and prosodic features relevant in differentiating response functions. However, there might also be more to gain from exploring how the phonetic/prosodic features of responses can be sensitive to the phonetic/prosodic features in the surrounding context. On the basis of the current study, I would argue for a more relational approach to the meaning of phonetics than has been offered so far, for phonetic research in general and for research on response tokens (and 'back-channels') in particular.

The phonetic characteristics of 'dts' correspond somewhat to those found by Curl, et al. (2006). They found that when speakers repeat their own previous turn in context of closing a sequence of talk, both productions have falling pitch contours while the repeat has lower pitch peak, shorter durations, and have similar loudness and articulatory characteristics in relation to the turn being repeated. The main differences between these findings and the current ones seem to be the loudness and articulatory characteristics. A possible reason for this is the difference between the two practices

studied. The current study dealt with how an uptake of talk is avoided ('dts'), whereas Curl, et al. (2006) studied a more explicit way of closing a sequence/topic. At the same time, both studies show that a repeated ('less') action is accompanied by phonetic characteristics that typically involve not being louder, not having higher pitch peak, and not having closer (consonants) and more peripheral (vowels) articulation.

The phonetic characteristics of 'Ndts' were typically 'opposite' of those for 'dts'. These characteristics correspond to some extent with those previously found for affiliating tokens (e.g. Müller, 1996), with more intonational variation than non-affiliating ones. But again, the current study suggests that the relational aspects of phonetic characteristics are also important, and also adds more phonetic detail to such analysis. Previous studies maintain the primary role of prosody/intonation in defining function (e.g. Ward, 2004), but there are no *a priori* reasons why pitch/intonation should be more important than other phonetic parameters.

There are several other points to pursue in further research. First, the binary distinction created here does not do full justice to the variety of action speakers are actually involved in. There could potentially be several sub-categories involved, which might account for some of the overlap in phonetic terms. One such issue is whether the exact nature, and phonetics, of the talk surrounding is of relevance to the production of the response tokens. Walker (2004a) investigated different types of increment, and found that the phonetics of increments depend on their relation to the previous turn unit (its 'host'). No obvious connection between type (and phonetics) of increment and the response tokens were discovered in the current analyses, but this might be worth further investigation. Also, regarding phonetic detail, a task for future research is to take a more integrated approach to phonetic detail, by focussing more on the combination of phonetic features rather than a set of single phonetic parameters. One could also include non-verbal information in future studies, for example by examining the combined use of verbal responses and head-nods, and whether or not they may complement each other in 'doing the same', or 'NOT doing the same'.

4.5.1 A final note in connection to the upcoming chapter

The use of head-nods has not been a primary concern in this chapter. However, several cases in the collection include head-nods, either as accompanying a hearer's verbal response, or used in similar sequences without any verbal response. In most cases these head-nods occur in the same turn-slot as the verbal responses studied above, i.e. turn unit 2 and 4. However, in a few instances, the head-nod continues from turn unit 2 and during the increment until turn unit 4, thus forms a continuous response. Example 4.10 above is such an example, where it was noted that Oscar continues nodding as part of displaying anticipation with Anne's action projection/turn production. Such uses of head-nods will be the central concern of the next chapter, where it will be shown how such nods take part in shaping the production of a turn, and that in particular the *timing* of these head-nods with a speaker's turn plays a crucial role in achieving trouble-free turn boundaries.

Whereas this chapter focuses mainly on hearer's actions, the next analytic chapters give more detailed attention to the dynamics of speaker *and* hearer contributions. Rather than investigating their contributions as separate, and organised sequentially in turn-slots, the next chapter focuses on speaker and hearer actions parallel and continuous activities, in negotiating towards a turn boundary.

CHAPTER 5

ANTICIPATORY NODDING

During a conversation it is necessary for speakers and hearers to maintain shared understanding of what the talk is about. For instance, it is on the basis of shared understanding, say regarding a person reference, that a speaker may proceed to talk about that person. Often shared understanding is established with the use of repair (Schegloff, 1992), or explicit formulations like *do you understand?*; but for most of the time a speaker assumes that a co-participant understands while they are talking, i.e. shared understanding is "assumed unless trouble is indicated" (Heritage, 2007, p. 259). In this chapter I will demonstrate one way in which such assumed (or implicit) shared understandings are not simply assumptions, but interactional achievements based on finely tuned and timed non-verbal behaviour, even within single turns of talk.

In the particular phenomenon studied the hearer nods in parallel with the speaker's turn. The hearer does this to display alignment and/or understanding with the current talk, but also to display anticipation of the rest of the speaker's turn (hence the title 'anticipatory nodding'), and in this way hearers facilitate shared understanding in further turn production. Thus the nodding adds something crucial to the talk in progress that would not be there otherwise, and as the analysis will show, the display and achievement of shared understanding is only successful if the hearer continues to nod throughout the speaker's turn. I will argue that the use and extension of head-nods with the ongoing turn defines what the interactional relevance of an upcoming transition place is (e.g. whether there is indication of trouble).

Example 5.1 below is used as a means to introduce this phenomenon, and as a starting point for the upcoming analyses. Here (lines 01/04) Lars explains how he manages well at a technical university, despite not having an engineering degree. This is done in response to Tor, who before argued that it is important if not necessary to have some engineering background in order to do research in this institution. Of main interest is

Tor's response following a pause in Lars' compound construction, following *den tekniske greia så...*/"that technical thing..." at the end of line 01.

For ease of access I have included only verbal content in extract 5.1, whereas a more detailed presentation of the target events, including head-nods, is represented in transcript 5.1a. In 5.1a, '//' stands for start and end of a head-nod group, '^' for head-nods, and '^v' represents distinctly larger head movements than elsewhere in the same nod group (for more information about transcription conventions and presentation, see chapter 3, section 3.4.3).

```
(5.1) KTH-NO, TL, 10:10/730 "nosebleed"

01    L:             °thh om jeg ikke (da/nå) har   hele den    her  `NOSEbleed
                         IF I    NOT  (part)  HAVE  WHOLE THIS  HERE  NOSEBLEED
                     °thh although I don't have this whole nosebleed

                     eh:: nh°  den  `TEKniske  greia så—
                             THAT  TECHNICAL STUFF part
                     uh:: nh°  that technical thing

02    T:             mm,
                     mm

03                   (--)

04    L:             `FUNker det likevel(m)—=            Tor nods (see transcript 5.1a)
                      WORKS   IT   STILL
                     it still works(m)

05    T:             =mm,
                      mm

06                   (1.5)

07    T              °pth ja  jˀ je:gˀ (-) jeg ´FÅR en assosiasjon til han eh:
                         YES I  I          I   GET AN ASSOCIATION TO  HIM
                     °pth yeah I- I (-) I get an association to (that one) uhm:



(5.1a) KTH-NO, TL, 10:10/730, "nosebleed" HEAD-NOD ANNOTATION

01    L:             den  `TEKniske  greia så—
                     that technical thing

02-06  HN(T)         //^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^/^v^v^v^v^v^v^v^//
02-06  T:            mm, (--0.5--)=                        =mm, (---1.1---)

04    L:                              =`FUNker det likevel(m)=
                                          it still works
```

Following Lars' 01, a displayed alignment (or disalignment) is a relevant next action from Tor, as Lars with his own experience contradicts the argument raised by Tor earlier, that it is necessary to have a technical/engineering background to study at the University Lars attends. In line 02 Tor displays such alignment with an *mm* and a nod. Tor then continues nodding during the gap of 0.5 seconds (03), and also while Lars proceeds on the second part the compound construction in 04 (*funker det likevel*/"it still works"). At the point where Lars reaches syntactic completion of his compound construction (04), Tor nods more vigorously than before accompanied by a second *mm* (05). Tor then proceeds with a next turn (07).

Tor displays an ability to anticipate where Lars' turn is heading, and collaboratively they establish the shared understanding that Tor aligns with Lars. This is done within the confines of a turn. The timing, extension and manner of Tor's response contribute to this achievement in three ways, or steps. First, by responding (*mm* + nod) in the middle of Lars' compound construction, Tor displays alignment with the present turn material while anticipating what the rest of the turn will be. Second, by nodding continuously throughout the rest of Lars' turn, Tor shows that no further turn material changes his alignment with Lars. Third, by nodding more intensively as Lars reaches turn completion (accompanied by a second *mm*), Tor demarcates both the turn and the action that promoted Tor's displayed alignment as complete. Shared understanding now appears confirmed, in the sense that there is nothing previous to this demarcation point that displayed a potential problem in establishing shared understanding, and the demarcation displays that this is still the case.

This sequence of events is representative of the sequence (and interactional process) explored in this chapter, the main objectives of which will be to:

- Demonstrate the interactional relevance of this sequence for the achievement of (implicit) shared understanding during the production of a turn.
- Demonstrate how this process defines turn boundaries in terms of what is relevant next.

The shape and description of the sequence in study is formalised in Table 5.A below. The three steps represent the key points that interactants orient to when mutually

treating shared understanding as having been achieved. The structural elements in this process will be referred to with the traditional category turn-constructional unit (TCU; see chapter 2), where 'mid-TCU pause' refers to the point where a speaker makes a halt in the production of the TCU, and 'mid-TCU response' refers to hearer's response to that mid-TCU pause.

Table 5.A. A formalised sequence of events for maintaining (implicit) shared understanding during the production of a turn of talk.

Step	Interactional process	Resources used by speaker and hearer
1	*Speaker: Action that makes a display of shared understanding relevant*	*Speaker:* A combination of linguistic/phonetic and non-verbal resources (mid-TCU pause)
2	*Speaker/hearer:Display and orient to the maintenance of shared understanding during the production of a turn*	*Speaker:* Continues on projected TCU **Hearer:** Nods in full co-extension with the TCU
3	*Hearer (and speaker): Display that the action that promoted shared understanding is complete*	*Hearer:* Uses verbal /non-verbal means to mark TCU completion, and to confirm that shared understanding is achieved

First, there is an action that promotes the display of hearer alignment and/or understanding, as an implicit display of shared understanding (step 1). We will see that the speaker may also use certain linguistic, phonetic or non-verbal cues to facilitate a hearer response in the middle of a TCU. Then there is a continued orientation to the maintenance of the (implicit) shared understanding (step 2). Finally the interactants establish that shared understanding was maintained all along, and talk may proceed to a next element (step 3).

As comparable cases are analysed throughout this chapter, the formalised sequence in Table 5.A will be used as a core reference point. The analysis starts with examples that

offer positive evidence for the sequence described above (section 5.3), i.e. examples that follow each step of the proposed sequence exactly. The analysis then proceeds with negative evidence (section 5.4), showing that when one of the proposed steps is violated, this will have interactional consequences regarding the maintenance of shared understanding. Finally (section 5.5), a set of deviant cases will show that (implicit) shared understanding can be achieved even though the formalised sequence is not followed as proposed. Rather than contradicting the proposed sequence, these examples will confirm the general claim regarding the collaborative nature of a turn production.

Before the analysis, some background will be provided on the study of head-nods and the process of shared understanding in talk-in-interaction (section 5.1). This is followed by a description of the dataset in relation to that of the previous chapter (section 5.2). The chapter ends with a summary and discussion of the findings (section 5.6).

5.1 Background

Head-nods are of interest to studies on talk-in-interaction because, just like verbal responses, they are resources with which interactants display among other acknowledgment, affiliation and agreement with the current talk (e.g. Maynard, 1987; Kendon, 2004; Stivers, 2008). In this section I will provide a background for studying nodding as part of the interactional management, and further situate this study among previous accounts regarding the maintenance of shared understanding.

5.1.1 Head-nods

Head-nods have previously been studied in terms of their role as back-channels, and as cues to turn-taking (Duncan, 1974; Maynard, 1987). Maynard (1987) studied the use of head-nods in Japanese turn-taking, and described their different pragmatic functions, e.g. as continuer (hearer); and defining clause boundaries, emphasis, affirmation, end of

turn and transition fillers (speaker). Maynard (1987) concluded that head-nods are accessible to both speaker and hearer, and that "head movement serves an important interactional function for turn negotiation at crucial moments of conversational exchange" (p. 600). However, this study does not go into any detailed analysis of how such negotiation works, and persists in viewing nodding as a "single behaviour" (p. 591), i.e. not as behaviour that is concurrent on emerging talk.

A study that is oriented towards the sequential relevancies of head-nods, and thus has more in common with the current study, is that of Stivers (2008). She focuses on nods in response to storytellings, and claims that verbal and nodding responses play different interactional roles. In the middle of the telling, verbal responses were found to align with the ongoing telling, whereas head-nods were found to claim access to the content of the telling, and to the teller's stance towards the telling. Mid-telling head-nods were thus conceptualised as affiliative with the telling/speaker, and verbal responses as structurally aligning, i.e. functioning more or less as continuers (Schegloff, 1982). Stivers (2008) stresses the relevance of sequential position of head-nods, and provides evidence for how speakers/hearers treat head-nods as ill-fitting or *non*-affiliative with the telling/speaker, when it is placed at the end of a telling.

There are in particular two observations in Stivers' (2008) paper that are interesting for the current analysis. The first one regards the role of mid-telling nodding as an early indication of affiliation. Stivers (2008) notes that the head-nods, by displaying access to the event told about and/or to the teller's stance, are "understood as forecasting a likely affiliative stance at story completion" (p. 53). Stivers (2008) does not make explicit whether or not a successful forecasting requires that the nodding co-extends with final parts of the storytelling, however this is apparently the case in at least some of Stivers' examples (e.g. example 6 on p. 41, where the nodding co-extends a third element in a chain of reported speech). In other words, Stivers' (2008) may in fact be addressing similar processes to those addressed in this chapter.

The second observation I would like to point out regards the ways in which a speaker may trigger a head-nod affiliation from a hearer. In Stivers (2008) it is made apparent (e.g. example 6 on p. 41) how a speaker may use gaze to call for hearer participation.

Heath (1992) also provides examples where a hearer's participation (with head-nods) are triggered by details in the speaker's verbal and non-verbal behaviour. These findings also correspond to studies by Goodwin (1979; 1981) and Hayashi (2003a).

These observations are relevant for my own analyses. However, rather than focussing on head-nods as single (Maynard, 1987) and sequentially sensitive items (Stivers, 2008), I will offer analyses that show how the exact temporal relations between nodding and ongoing talk is important for the interactional management, specifically in terms of maintaining shared understanding during the production of a turn. Note that I will consistently use the term alignment (and understanding) when describing hearers' actions in the upcoming examples. There is some basis in the literature for distinguishing between alignment and affiliation (e.g. Stivers, 2008), however I have found it appropriate to use only the former term in this study.

5.1.2 Shared understanding

The concept of shared understanding has previously been addressed by CA researchers, under such terms as 'shared cognition' and 'intersubjectivity' (Schegloff, 1991; Heritage, 2007), and by other communication researchers, using terms like 'common ground' (Clark & Brennan, 1991). These authors essentially describe the same thing, but I intend to use the CA-based research as the basis for my analyses, since its technical and methodological underpinnings has more in common with my own research.

The main issue regarding shared understanding in CA is that it touches upon the interface between the structures of interaction, and cognition, i.e. meaning-making, and according to Schegloff (1991), CA provides the methods for studying *how* this works. He states that: "Practices of conduct in ordinary interaction can be examined for the ways in which they furnish or embody procedures by which a sense of a world known in common is reinforced and implemented" (p. 153). CA has identified a few particular 'practices of conduct', including repair (Schegloff, 1992), and person reference (Heritage, 2007). The repair studies highlight that this practice is a *defence* of intersubjectivity, which can occur both within (i.e. self-repair) and in one of the turns

following the turn that gets repaired. In his study on person (and place) reference, Heritage (2007) highlights that because it is too costly to constantly check a co-participant's recognition of a person explicitly, a certain balance between progressivity and intersubjectivity needs to be maintained. Heritage (2007) claims that reference-making is strongly biased towards progressivity, but that "it ultimately rests entirely on the hidden work that speakers do to ensure that their references to persons are recognizable without the need for repair" (p. 279). Notice that Heritage (2007) refers to speakers' 'hidden work'. This is not a very precise description, and it implies that participants' assumptions regarding shared understanding are merely assumptions, i.e. not based on observable interactional work (e.g. bodily behaviour). Although it is not investigating person references in particular, the current analyses attempt to pin down and demonstrate what such hidden work might be.

5.2 Notes on data collection

The examples used in this chapter were collected as a subset of the collection used in the previous chapter. The two datasets are similar in that they address a sequence of short verbal responses from a hearer, in response to emerging elements in a speaker's turn. The two datasets are different in terms of structural placement of the responses, and their interactional relevance.

In terms of structure:

- The initial hearer response in the current study occurs within one TCU (i.e. at a place of non-possible completion), whereas it occurs between a complete TCU (host) and an increment in the previous chapter (i.e. at places of possible completion).

These structural differences are reflected in the interaction:

- Although there were head-nods found in the dataset in the previous chapter, these were generally not produced in a continuous fashion (i.e. along with the speaker's increment). This suggests that participants themselves treat TCU host

+ increment as separate items in a sequence of talk, whereas the structure in the current study (mid-TCU pause + TCU completion) is treated as one, collaboratively shaped, unit.

- Further, the mid-TCU responses are facilitated by the speaker's turn design (mid-TCU pause), rather than structurally provided for by the end of a TCU. Thus, the speaker is more actively seeking displayed understanding from hearer in the current dataset, than in the dataset of the previous chapter.

No particular activity sequence was selected (e.g. end of telling, insert sequence); the main focus was the phenomenon of anticipatory nodding, and interactants' orientations to shared understanding during the production of a turn/TCU. Sequential differences will be considered and accounted for as the cases are presented individually.

The main criterion for data selection was that there would be a mid-TCU pause followed by hearer response. There was no pre-definition of what a mid-TCU pause would look like, apart from being a pause somewhere after a TCU beginning and prior to a TCU completion, and that it did not have any design features of a turn completion. That is, given that the mid-TCU pause was incomplete in terms of syntax, there would be no audible or visible indication that the incomplete syntactic unit would still be designed as complete in action terms (cf. chapter 4, section 4.2.1, for further descriptions of what may constitute completeness in terms of speech production).

The examples will be presented according to the conventions presented in chapter 3 (section 3.4.3), in a case-by-case manner. No quantitative data are made available for this study, the claims are supported entirely through qualitative and observational means.

5.3 Positive evidence: Inviting and securing shared understanding

The analysis starts with examples that offer positive evidence to the relevance of anticipatory nodding for achieving shared understanding. The first subsection (5.3.1) will focus on how a mid-TCU pause makes relevant the hearer's displayed alignment (i.e. step 1 in the proposed sequence in Table 5.A), whereas the second subsection (5.3.2) will focus on how the interactants maintain shared understanding (step 2) and mark turn completion, and thereby confirms that shared understanding is achieved (step 3).

As previously, I will present only the verbal content first, followed by the relevant part of the transcript including non-verbal detail. I have indicated the 3 steps from the schematised sequence (Table 5.A) in both transcripts. In the verbal-only transcript, which will be referred to as 'the main transcript', they are placed to the left of the relevant transcription lines. In the second, non-verbal transcript the steps are indicated at their precise moment of occurrence.

5.3.1 Displayed understanding triggered by a mid-TCU pause

Two instances, presented in transcripts 5.2 and 5.3, will be presented to illustrate how a hearer's display of understanding is both interactionally relevant and negotiated towards with the use of a range of linguistic/phonetic elements, and gaze. The first example, 5.2, will show how mid-TCU pauses can trigger a displayed understanding even when the interactional relevance for that display is not very clear. The second example, 5.3, will show how particular phonetic features in a mid-TCU pause can also trigger a clearly relevant display of understanding.

In both examples Lars is talking about his trip to Athens, where he met with a friend who is a local to the city (example 5.3 follows immediately after example 5.2). Prior to the excerpt in 5.2 Bengt has requested Lars to talk about the things Lars and his friend

saw and did there (i.e. Bengt asks: *var det kult da*/"was that good"). The transcript starts with Lars initiating a multi-unit response to that request in 01.

```
(5.2) KTH-NO, BL, 04:20 "aleine"

01    L:            °pthh jo  `DET ↑var `KULT.
                        (YES) THAT WAS  COOL
                    °pthh yeah that was good

02    L:            og  jeg fikk jo   `SE  saker  som
                    AND I   GOT  part SEE THINGS THAT
                    and I got to see  things that

      1->           jeg ikke skulle ha  ¯SETT:— (-) om je:g— (eh)=
                    I   NOT  SHOULD HAVE  SEEN     IF I
                    I wouldn't have seen:     (-) if I: (uh)

03    B:    2->     =mm,
                     mm

04    L:    2->     hadde gått a`LEI´ne?
                    HAD   GONE ALONE
                    had gone on my own

05    B:    3->             [mm;    ]
                             mm

06    L:            <<all >[for ekse]mp>el: (eh) om du  har ´HØRT  om
                           FOR EXAMPLE          IF YOU HAVE HEARD ABOUT
                           for example (uh) have you heard about those

                    de  `OPPtøyene som  var i:
                    THOSE RIOTS   THAT WERE IN
                    riots (that were) in:
```

The construction *som jeg ikke skulle ha sett om jeg...*/"that I wouldn't have seen if I..." in 02 clearly projects ...*hadde vært aleine*/"...had been on my own" in 04, given that Bengt and Lars already have established shared knowledge that Lars was accompanied by a local. With his response, *mm* and a nod in 03, Bengt (in addition to displaying hearership) shows that accessibility to the projected meaning is shared, and that he is able to anticipate where the turn is heading. However, Bengt's display of anticipation does not occur just anywhere, as we will see Lars' use of prolongations and halts, in combination with use of gaze, are central in inviting, and triggering, Bengt's response.

Head-nods and gaze are included in transcript 5.2a below: Gaze is annotated above the line for head-nods (indicated by Gz). In the gaze transcript 'x___' represents mutual gaze, while ', ,' represents gaze-shift, either away from mutual gaze, or towards mutual

119

gaze. Note that only the speaker's (in this case Lars') gaze is given. Unless otherwise specified the hearer (in this case Bengt) gazes back at the speaker (cf. section 3.4.3).

```
(5.2a) KTH-NO, BL, 04:20 "aleine" HEAD-NOD AND GAZE ANNOTATION

"STEP"                              1
02     Gz(L)                  , , , , , x_____
02     L:          jeg ikke skulle ha ¯SETT:— (0.3) om je:g— (eh)
                   I wouldn't have seen:    (0.3) if I: (uh)

"STEP"             2
03-04  Gz(L)            _____
03-04  HN(B)       //^^^^^^^^^^^^^^^^^^^^^^^^^^
03     B:           mm, =

04     L:                   = hadde gått a`LEIne?
                              had gone on my own
```

Notice that mutual gaze is established just prior to Lars' halt following *sett*/"seen" (line 02) and onwards. Gaze is systematically used to enhance the relevance for some kind of co-participation (Kendon, 1977; Goodwin, 1981), and seems to be oriented towards in such a way also here. By gazing at him, Lars orients to Bengt's online access to understanding. Bengt's response however, is only initiated after a second mid-TCU pause, following *om jeg*/"if I" (line 02). A potential account for this is that it might not be entirely clear why Bengt's displayed understanding is relevant at the point of the first pause, neither for the analyst nor for Bengt. But then, as Lars produces another mid-TCU pause, and maintains mutual gaze, this further *increases* the relevance of Bengt's displayed understanding.

Another relevant observation about the relevance of a mid-TCU pause in this example, is that Lars orients to Bengt's display of understanding, as an opportunity for him to immediately head on to a next turn, in line 06. In other words, mid-TCU pauses seem like an important interactional resource for securing shared understanding and in facilitating efficient turn-transition.

In further support of the above observations, there is no indication that Lars projects a next speech element at the mid-TCU pause, for example in the form of glottalisation or co-articulation/preparation for a next sound (cf. Local & Kelly, 1986). That is, although Lars' turn by no means complete in terms of syntax, the production of *om jeg* indicates

that although a syntactic completion may be relevant, it will not come right away. Lars keeps his mouth open during the pause, suggesting readiness to continue talking, however by not preparing for a next sound Lars does not prevent Bengt from contributing, but rather invites Bengt to do so.

Example 5.3 below further demonstrates the interactional relevance of a mid-TCU pause, providing even more powerful evidence for how a mid-TCU pause may trigger a hearer response. In this example there is an additional element in Lars' speech production which, in along with mutual gaze, strongly contextualises his halts and prolongations as projective of a mid-TCU response. This contextualising element lies in how Lars demonstrably 'reverses' the anticipation of a next sound before the mid-TCU pause, which Bengt then picks up on.

```
(5.3) KTH-NO, BL, 04:27 "femtenåringen"

06    L:           <<all >[for ekse]mp>el: (eh) om du  har ´HØRT  om
                       FOR EXAMPLE          IF YOU HAVE HEARD ABOUT
                    for example (uh) have you heard about those

                   de  `OPPtøyene som  var  i:
                   THOSE RIOTS     THAT WERE IN
                   riots (that were) in:
07    B:           ja,
                   YES
                   yes
08    L:    1->    °hhh hun `VISte  meg den  plassen han `FEMtenåringen:(eh)
                       SHE  SHOWED ME  THAT PLACE   HE   FIFTEEN-YEAR-OLDdet
                   °h she showed me the place where that fifteen-year-old (uh)

            2->    (.) ble skutt av poli`TI´et,
                       WAS SHOT  BY POLICEdet
                   (.) was shot by the police

09    3*->   (-)

10    L:           °hhh og  ´RUNDT: den  plassen så   var det ((...))
                       AND  AROUND THAT PLACE   part WAS IT
                   °hhh and around that place there was ((...))
```
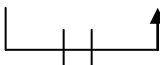
Here Lars continues the initiated talk about what he saw in Athens, referring to what he saw at the place where a young person was shot by the police, which happened in connection to recent riots in the city. Bengt initiates a mid-TCU response following *femtenåringen*/"fifteen-year-old" in 08, displaying his recognition of this shooting

incidence. This time Bengt displays his understanding with a head-nod and no verbal response. The first line is given as 06 since it continues immediately on example 5.2 above. Step 3 is marked with a '*' because this is a boundary example, where it is not entirely clear whether or not turn completion is marked (also not for Lars). This deviant feature will be further described in section 5.5.

Shared understanding becomes relevant as Lars makes a reference to a specific incidence during a chain of riot in Athens. In 07 Bengt displays recognition of these events with a *ja*/"yes" (notice that this is done before Lars completes the final phrase in the interrogative, *i (Athen)*/"in (Athens)"). This is a typical pre-sequence, which establishes that the audience is ready for the continuation of the telling (cf. Terasaki, 1976). As Lars proceeds with what appears to be the main part of the telling in 08, a second issue regarding shared knowledge emerges. That is, as Lars refers to the shooting of the fifteen-year-old in 08, it is not yet established whether Bengt also knows about this specific incident. As Lars produces *femtenåringen* he uses the definite article (–en)/"the", followed by a pause. This definite article confirms that Lars is referring to one particular boy, and arguably the relevance for shared understanding has reached a peak at this moment. And as we will see, Bengt also treats this as a moment for him to display recognition of the referent.

Transcript 5.3a will be used to illustrate how the mid-TCU pause triggers Bengt's displayed understanding. In addition to a representation of Bengt's head-nods, the transcript includes a phonetic transcription of Lars' production of *femtenåringen*, and a broken arrow signifying reversed co-articulation.

```
(5.3a) KTH-NO, BL, 04:27 "femtenåringen" HEAD-NOD AND PHONETIC ANNOTATION

STEP                  1                2
08    HN(B)                            //^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^
08    PHON(L)          [fæmt nɔ ɾɪŋəmn ə]
08    L:               `FEMtenåringen:(eh) (.) ble skutt av poli`TIet,

                       (the) fifteen-year-old (uh)(.) was shot by the police
```

Lars' production of *femtenåringen* is not a typical realisation of this word, as the realisation of the definite article (the word-final /en/) is realised with two nasal segments, one bilabial [m] and the other alveolar [n]. Both of these nasals are expected as a realisation of a definite article in Norwegian. That is, the definite article /en/ is commonly produced as a bilabial [m], especially when the following consonant is also bilabial, and /en/ is commonly produced as alveolar [n] when prior to other alveolar consonants, and in utterance-final position for example. It is the fact that Lars *changes* from bilabial to alveolar nasal that, from a linguistic perspective, is unexpected. But this production proves highly informative if we connect it to Lars' mid-TCU halt. Assuming that Lars is already heading for the bilabial in *ble*/"was" (indicated by an arrow), a change from [m] to [n] would put this co-articulatory action 'in reverse', and thereby make *femtenåringen* heard as final rather than immediately projective of more speech.

As in example 5.2 then, the mid-TCU pause is accompanied with an articulatory posture that does not project a specific next speech element. Example 5.3 further demonstrates that the interactants orient to the relevance of such a non-projective articulatory posture. Evidence for Bengt's orientation to this relevance is offered in how he initiates his head-nod at the exact moment following where [m] is changed to an [n]. More precisely, the nodding starts approximately one-tenth of a second after the nasal becomes alveolar. In other words, although his displayed understanding is in any case relevant (Lars has reached the definite article of a referent which Bengt might not be familiar with), Lars' co-articulatory reverse seems to trigger Bengt's response to occur exactly at this point in time.

In summary, examples 5.2 and 5.3 show how a speaker's turn design may actively trigger a hearer's mid-TCU response with the use of particular linguistic and phonetic resources, along with mutual gaze.

5.3.2 Maintaining and achieving shared understanding

The focus now is on the maintenance of shared understanding throughout and following the TCU completion, i.e. the management of steps 2 and 3 in the proposed

sequence in Table 5.A. Re-using example 5.1 and 5.2 above, the presented analysis will highlight how the use of anticipatory nodding facilitates turn completion, while displaying shared understanding, which is then confirmed at TCU completion with a differentiated response.

As we saw in example 5.1 in the introduction of this chapter, Tor's displayed understanding is relevant as Lars contests Tor's argument made earlier. Details, including Tor's gaze, are given in transcript 5.1b below.

```
(5.1b) KTH-NO, TL, 10:10/730 "nosebleed" HEAD-NOD AND GAZE ANNOTATION 2

((In 01/04 Lars explains how he manages well at a technical university, despite
not having an engineering degree. This is done in response to Tor, who prior to
the excerpt argued that it is important if not necessary to have some
engineering background in order to do research in this institution))

01    L:              den `TEKniske  greia så—
                      that technical thing

STEP                  2                                    3
02-06  HN(T)          //^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^/^v^v^v^v^v^v^//
02-06  HN(L)             //^^^^^^^//                      //^^^^^^^^^^^^//
02-06  Gz(T)          _____, , ,DL
02-06  T:             mm, (--0.5--)=                        =mm, (---1.1---) (0.4)

04     L:                             =`FUNker det likevel(m)=
                                       it still works

07     Gz(T)          DL
07     T:             °pth ja  j? je:g? (-) jeg ´FÅR en assosiasjon til han eh:
                      °pth yeah I- I  (-) I get an association to (that one) uhm:
```

This example shows how the interactants collaboratively display the achievement of shared understanding, on the basis of anticipatory nodding. Two parallel activities demonstrate this. First, by nodding throughout Lars' turn Tor contextualises his alignment as contingent on Lars' talk. Second, Lars actively uses this as an environment to accept Tor's alignment, (i) by nodding in response to Tor's (first) *mm* + nod, and (ii) by completing his compound construction. Furthermore, by intensifying his nodding at the end of Lars' turn, Tor marks that this is the point at which he does not expect any further turn development from Lars, i.e. he confirms that Lars' turn and their shared

understanding is for all practical purposes achieved. In response to Tor's intensified nod and an *mm*, Lars nods a second time, as if ratifying their achievement.

Notice the use of the particle *så* (01) in this example. This particle is also found elsewhere in the data, prior to a mid-TCU pause. In example 5.1 there are no clear prolongations in the production of this particle, nor a silence portion preceding Bengt's response. Still, this particle is oriented to as the point where a mid-TCU response is relevant, and it is possible that the *så* is a highly recognisable mid-TCU item in Norwegian, for which prolongations are not needed. However, the other accompanying features are intact, i.e. there is mutual gaze, and the phonetic production of *så* neither indicates turn-finality, nor projection of a next speech sound.

Below (transcript 5.2b) example 5.2 is re-visited as a second, slightly different illustration of this process. In this example step 3 takes a different form than in example 5.1. As in example 5.1, the hearer (Bengt) produces a second verbal response (*mm*) to mark turn completion. However, unlike 5.1, the TCU-final *mm* is not accompanied by a differentiated head-nod.

```
(5.2b) KTH-NO, BL, 04:20, "aleine" HEAD-NOD AND GAZE ANNOTATION

02     Gz(L)              , , , , , x_____
02     L:       jeg ikke skulle ha ‾SEtt:— (0.3) om je:g— (eh)
                I   wouldn't have    seen: (0.3) if I: (uh)

STEP             2                          3
03-06  Gz(L)    _____, , , DR
03-06  HN(B)    //^^^^^^^^^^^^^^^^^^^^^^^^^^^//
03-05  B:        mm, =                      [mm;            ]

04-06  L:               = hadde gått a`LEIne? [<<all >for ekse]mp>el: (eh)
                          had   gone on my own for example (uh)
```

This difference suggests that a differentiated head-nod at the end of the TCU is not essential as a step 3 achievement, as long as there is some other form of confirmation (e.g. verbal). Alternatively, the lack of head-nod in this case might be related to the fact that Lars continues speaking *(for eksempel/*"for example" in 06) at the same time as Bengt produces his TCU-final response. As further support for this claim, Lars shifts his

gaze away from Bengt as soon as his TCU is complete, and this way initiates and contextualises *for eksempel*/"for example" (06) as a transition away from the matters of the previous turn. Furthermore, Lars produces *for eksempel* faster than previous talk, and in this way quickly secures the hearable initiation of a next turn. In sum, Lars shows that he does not need any further confirmation of shared understanding, and Bengt displays his orientation to this as he stops nodding simultaneously with Lars' gaze-shift and next turn-initiation (I was not able to determine whether Lars' gaze-shift or Bengt's nod-stop occurred first).

In context of the preceding chapter on phonetic resources for 'doing the same', it is interesting to note that the phonetic relationship between the two *mm*'s in both 5.1 and 5.2, corresponds to the phonetic characteristics for 'doing the same' (see chapter 4). Also, observations about what happens next confirm the findings on what the interactional relevance of 'doing the same' is. First, in example 5.1 Tor proceeds on a next turn (07), tangentially related to Lars' talk (which he projects with a gaze-shift during the gap in 06). This observation fits with the claim in chapter 4 about how similar responses are used in connection with a topic shift. In this way, at the interface between the current study and the previous one, Tor might be doing two things at the same time: (i) displaying shared understanding, and (ii) not projecting further on-topic engagement. Correspondingly, in example 5.2 Lars projects a next turn as Bengt's understanding is secured and he does not project an on-topic engagement (in fact, he hardly gets the opportunity to).

5.3.3 Summary

This section has demonstrated how speakers use mid-TCU pauses to attract hearer alignment/understanding, and how shared understanding is displayed and confirmed during and after a turn production.

In context of an incomplete TCU, a range of resources are relevant in inviting a mid-TCU response, including:

- Prolonged speech sounds, followed by a pause

- Mutual gaze

- Linking particles like *så*

- Definite article (which in Norwegian occurs word-finally)

- Phonetically *not* projecting a next speech sound, i.e. no glottalisation or coarticulatory features prior to mid-TCU pause

Hearers display their understanding following the mid-TCU pause and throughout the rest of the TCU, which constitutes shared understanding. The achievement of shared understanding is confirmed by providing a differentiated head-nod, and/or a verbal response. The confirmation is sensitive to the emerging interaction: Nodding does not continue any longer than until shared understanding is confirmed.

5.4 Negative evidence for the interactants' orientations to anticipatory nodding

To further prove the relevance of the sequence in Table 5.A, it is necessary to establish whether the observed patterns do indeed make a difference for the maintenance of shared understanding. For example, does it make a difference for the interactants whether the mid-TCU pauses are responded to, whether or not the nodding co-extends with the TCU-completion, and whether or not the TCU-completion is marked? Based on Table 5.A one could expect that:

- When there is no anticipatory nodding following a mid-TCU pause (violation of step 1), the further production of the turn breaks down, and is sought to be resolved with a repair-initiation

- When there is nodding, but the nodding stops prior to TCU completion (violation of step 2), this displays a problem in maintaining shared understanding

- When nodding extends throughout the TCU, but continues in an undifferentiated manner after the TCU is complete (violation of step 3), this displays that the action that promoted shared understanding is not yet finished.

No observations were made of step 1 violation. This may in itself be evidence that when mid-TCU pauses occur, they are oriented to as making a displayed understanding relevant. In example 5.2 above two pauses were needed for a hearer to respond, and shows that the further production of a turn might be affected by a missing response. However, as the response does eventually occur and the turn production does not break down, this does not qualify as a violation of step 1. Such examples might be revealed in future analysis.

On the other hand, examples where steps 2 and 3 are violated were found in the data, and will be presented in two subsections below.

5.4.1 When nodding does not co-extend with a TCU completion

First is an example (5.4) that shows that when the nodding is stopped prior to TCU completion (violating step nr. 2), it is oriented to as a failure in shared understanding. Prior to this excerpt Lars has been explaining how he did the experiments for his PhD research, which addresses musical scratching. In these experiments he focussed on the movements used to scratch rather than studying more musical elements in scratching. Tor seems to have come to the understanding that the experiments were done without considering or using any musical context, and in lines 03-08 he argues that it is not possible to study scratching by isolating the scratching from the music. However, Lars treats Tor's displayed understanding as unfitted with his own conceptions, and this is most clearly demonstrated in the interaction following line 08. The mid-TCU pause is found at the end of line 06 (step 1). Step 2/3 is replaced by a '*', indicating a problem in maintaining shared understanding.

In this example the participants are visibly orienting towards the turntable present in the recording studio. The turntable works as the point of reference for talking about musical scratching (see Appendix C for a further description about 'scratching' and 'turntable').

(5.4) KTH-NO, TL, 15:29 "musikken i bakgrunnen"

```
01    T:              for    når  du  gjør eksperi´MENTene så    må
                      BECAUSE WHEN YOU DO   EXPERIMENTSdet part MUST
                      because when you do the experiments you have to

02                    du  må   jo  `HA  ´inn:: (nh°) (.)
                      YOU MUST part HAVE IN
                      you have to get/record (.)

03                    det går  jo   ikke å `GJØre det sepa´RAT   `mener jeg;
                      IT  GOES part NOT TO DO   IT  SEPARATELY  MEAN  I
                      it's not possible to do it separately I mean

04                    iso´LERT  `sånn.=
                      ISOLATED  LIKE
                      (like) isolated

05    L:              =[nei-        ]
                        NO
                        no

06    T:              =[(liksom det)] det er jo: det må  jo   være
                        (LIKE)    IT   IT IS part IT MUST part BE
                        (you know it) it is it has to be

      1->             rela^TERT til den: (-)
                      RELATED   TO  THAT
                      related   to the (-)

07    L:    2->       [mm,=
                       mm

08    T:    *->       =den  mu´SIKken man spiller i  `BAKgrun[nen   eller-]
                       THAT MUSIC      ONE PLAYS  IN BACKGROUNDdet OR
                       the music you play in the background or

09          *->                                          [°pthhh      ]
                                                          °pthhh

10    T:    *->       °hhh   eller hvordan:-=
                             OR    HOW
                      °hhh   or    how

11    L:    3*->      =´jo`da.
                        YES
                        yes

12                    (.)

13    L:              men eh: [d:eˀ ]                          [d::  ]=
                      BUT      THEY                             (THEY)
                      but uh:  the (ones)                       (th::)

14    T:                      [eller] <<all >fordi   fordi>  du d[: f f]=
                               OR            BECAUSE BECAUSE  YOU
                               or     because because         you    ( )

15    L:              =de  jeg har [`SPILT] inn de   de   har  fått eh:m:-
                      THEY I   HAVE PLAYED  IN  THEY THEY HAVE GOT
                      the ones I've recorded they've got uh:m:

16    T:                          [mm?   ]
                                   mm
```

Although Lars conforms to Tor's candidate understanding with a *joda*/"yes" in 11, it appears clear for both participants that Tor's understanding calls for modification. Lars initiates a *men eh*/"but uh" in 13, followed/overlapped by Tor's attempt to modify his own point in 14 (*fordi fordi*/"because because"), which is then abandoned in favour of Lars in 15, where he starts from scratch trying to explain how the experiments were done. Apparently, the participants in Lars' experiments were given the same samples, or pre-sets, to scratch on. In other words, the experiments were not done separately from musical context, but with a simplified musical context. This accounts for how it was problematic for Lars to support Tor's candidate understanding (and also to straightforwardly disagree).

These attempts at fixing Tor's candidate understanding all start following Tor's mid-TCU pause at the end of line 06 (Tor's *relatert til den*/"related to that"), when Lars stops his anticipatory nodding prior to Tor's TCU completion in lines 08-09. Thus, the nod-stop is the first indication of disalignment. Details are given in transcript 5.4a below.

```
(5.4a) KTH-NO, TL, 15:29 "musikken i bakgrunnen" HEAD-NOD AND GAZE ANNOTATION

STEP                 1                 2
                                       {turntable}
06-08  Gz(T)    _____, ,{   } , , , x_____
06-08  HN(L)                      //^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^
06-08  T:       rela^TERT til den: (-)=     =den  mu´SIKken man spiller i
                related    to the (-)        the  music      you play   in

07     L:                                = mm =

STEP                 (2)      *
08-11  Gz(T)    _____
08-11  Gz(L)    _____, , R         , , {turntable}
08-11  HN(L)    ^^^^^^^/                                  /{head-thrust}
08-10  T:       `BAKgrun[nen   eller-] °hh[h   eller hvordan:-=
                the background or       °hhh    or    how

09-11  L:                 [°pthhh    ]                    =´jo`da.
                           °pthhh                            yes
```

Tor's talk is related to the turntable present in the room, which explains his gaze towards this object in the middle of his turn in 08 (following *relatert til den*). This gaze

shift may project the next item in his turn, a sense which Lars then relevantly displays by producing an *mm*.

Right before Tor reaches the end of his turn, during *bakgrunnen*/"the background" (08), Lars stops nodding and initiates an inbreath. This, and the absence of a step 3 response, displays Lars' disalignment, and the following evidence suggests that Tor early on treats Lars' behaviour as such: During Lars' inbreath Tor initiates his first attempt to offer a modifying stance towards his own argument, with an *eller*/"or" (end of 08). Tor then continues, quite disfluently, to display orientation to Lars' disalignment with *eller hvordan*/"or how" in 10. Lars, in parallel with Tor's *eller hvordan*, displays his efforts at resolving shared understanding by gazing at the turntable.

Thus, what is missing in this example compared to the previous examples, is participants' orientation towards their parallel behaviours as constituting shared understanding (following 08). One can tell that this is missing because the participants continue to work towards an understanding. I argue that Lars' discontinued head-nod and missing verbal response are treated as a display of such disalignment in this example.

5.4.2 No marking of TCU completion

Two examples will demonstrate how no marking of TCU completion (violation of step 3) makes relevant further contributions to shared understanding, i.e. that the action that called for hearer alignment/understanding is not complete for all purposes.

The first example (5.5) addresses the hearer's response to a story-telling. The target turn occurs at the final part of a story-telling, i.e. where it approaches its climax. At such a point it is relevant for a recipient to show appreciation and understanding of the story, or identify its main point as funny, interesting, terrible, etc., depending on what the design of the story projects (Jefferson, 1978). In other words, a more elaborate uptake than a head-nod is relevant in the case of a story-telling (cf. Stivers, 2008), compared to the examples presented above. Corresponding to previous examples, anticipatory nodding displays an early appreciation/understanding of the telling. But the nodding continues in an undifferentiated manner beyond the relevant TCU-completion (the climax point), which is oriented to as maintaining the relevance of an appropriate understanding, i.e. the hearer *avoids* displaying that shared understanding is fully achieved. This is demonstrated by the choices Bengt and Lars make following the climax of the story.

Lars talks about an incident with a taxi ride in Athens, where the taxi driver did not give priority to an ambulance. The reason for telling this story is that Lars has experienced Greeks as being a bit rude and not very compassionate.

```
(5.5) KTH-NO, BL, 09:12 "taxi"

01    L:            °hh vi kjørte taxi(m)
                        WE DROVE  TAXI
                    °hh we took a cab

02    L:            °ptk og  det   var ganske mye  trafikk
                        AND THERE WAS PRETTY MUCH TRAFFIC
                    °ptk and there was quite a bit of traffic

03                  (-)/((B: nod))

04    L:            °mh så   kom  det en ambulanse bak   oss
                        part CAME IT  AN AMBULANCE BEHIND US
                    °mh then an ambulance came behind us

05                  (-)/((B: nod))
```

```
06    L:              eh det   var  fire filer eller noe  sånt tre   filer(m)
                        THERE WERE FOUR LANES OR    SOME SUCH THREE LANES
                      uh there were four lanes or something three lanes

07                    (-)/((B: nod))

08    L:              °pth (så)   kom det   en ambulanse i  utrykning bak    oss
                         (part) CAME THERE AN AMBULANCE IN EMERGENCY BEHIND US
                      °pth (then) an ambulance came       in emergency behind us

09                    (-)/((B: nod))

10    L:              °mh men ettersom eh det   fantes en liten luke: h°
                         BUT SINCE      THERE WAS   A  SMALL POCKET
                      °mh but since uh there was a small pocket h°

11    L:     1->      der han kunne kjøre inn taxien s[å: b kjørte han der]=
                      THERE HE COULD DRIVE IN TAXIdet part  DROVE  HE  THERE
                      where he could enter the taxi        he drove there

12    B:     2->                                   [nhh° nh°      nh°  ]=
                                                    ((laughter/nod))

13    L:     2->      =[rett  og  la  seg foran den [her] t (før) ambulansen   da
                       STRAIGHT AND PUT ref. AHEAD THAT HERE     AMBULANCEdet THEN
                       straight and put himself in front of the the ambulance

14    B:     2->      =[nh°                        [°h ]
                        ((laughter/nod))

15           *->      (---)/((B: nod))

16    B:     *->      [o:g eh:: den  står    der  ] og  tuter=
                      AND       THAT STANDS  THERE   AND HOOTS
                      and uh that one stands there and hoots

17    L:     *->      [((headshake/palms up))      ]

18    B:             =og [han] sier et  eller annet (.) kjipt på (.)
                       AND HE  SAYS ONE OR   OTHER    LAME ON
                      and he   says something    (.) lame on  (.)

19    L:                 [ja ]
                          YES
                          yes
```

The turns in focus are in lines 11-16. In lines 10-11 Lars approaches a potential climax in his telling, and as he produces the mid-TCU particle *så* (11), Bengt initiates his response. Bengt's response (12/14) is made of nasal outbreaths (small laughter tokens) and a continued nod. The outbreaths continue during most of Lars' further turn construction (until his inbreath aligned with Lars' *her*/"here" in 13), while the nods continue throughout Lars' turn (13), and beyond it (15), in an undifferentiated manner. See transcript 5.6a for details (including Bengt's smiling gesture; mutual gaze is maintained until Bengt's turn in 16).

```
(5.5a) KTH-NO, BL, 09:12, "taxi" HEAD-NOD AND FACE ANNOTATION


11      L:              der han kunne kjøre inn taxien s[å: b kjørte han der]=
                        where he could enter   the taxi      he drove there

        STEP            1                              2
12      HN(B)                                          //^^^^^^^^^^^^^^^^^^^
12      Fc(B)                                          //smile-------------
12      B:                                             [nhh° nh°      nh°  ]=

        STEP            (2)
13-14   HN(B)           ^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^
13-14   Fc(B)           --------------------------------------------------------
13      L:              =[rett og  la seg  foran den [her] t (før) ambulansen   da]
                          straight and put himself in front of the the ambulance

14      B:              =[nh°                           [°h ]

        STEP            *
15-16   Gz(B)           _____, , DR
15-16   HN(B)           ^^^^^^^//
15-16   Fc(B)           -------//
15-16   B:              (-0.9-) [o:g eh:: den står der  ] og tuter
                                and uh that one stands there and hoots

17      L:                      [((headshake/palms up))]
```

Bengt's undifferentiated nodding (15) displays an orientation to the continued
relevance of establishing shared understanding. One could ask whether a differentiated
nod (and/or a short verbal response) would be appropriate in this case. Following
Stivers (2008) it is not, and Bengt's continued nod beyond the TCU is then a resource to
show that although an appreciation is not available at the moment it might still be
coming up. Both interactants orient to this relevance: Lars offers his own stance toward
the telling with a facial-bodily gesture in 17, involving a head-shake and a palms-up
gesture. This gesture seemingly displays a lack of further words or comments available
(i.e. "what can you say"). In overlap (16), Bengt initiates a further elaboration of the
telling; apparently he seeks to draw a more complete picture of the situation, and
thereby orienting to his own response so far as insufficient as response to Lars' telling.

Example 5.3, here re-presented as transcript 5.3b (including gaze and head-nods), also
shows that when step 3 is violated, the relevance of promoting shared understanding
continues beyond the end of the relevant TCU. Bengt does not clearly confirm shared
understanding of Lars' reference (i.e. of the episode where a fifteen-year-old was shot
by the police) at TCU-completion. Bengt's undifferentiated response (and continued

hearership) orients to the continued relevance of Lars' telling, but at the same time Lars monitors Bengt for a confirmation before proceeding on his telling.

```
(5.3b) KTH-NO, BL, 04:27, "femtenåringen" HEAD-NOD AND GAZE ANNOTATION

       STEP          1               2
08     Gz(L)         _____
08     HN(B)                         //^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^^
08     L:            `FEMtenåringen:(eh) (.) ble skutt av poli`Ti´et,
                      fifteen-year-old (uh)(.) was shot by the police


       STEP          3?
09-10  Gz(L)         _____, , D
09-10  HN(B)         ^^(faster)^^^^^^^//
09-10                (--0.5--) °hhh og ´RUNdt: den  plassen
                               °hhh and around that place
```

A '?' is put next to step 3, because it is not clear whether or not Bengt visibly marks Lars' TCU completion (08). The video suggests that Bengt nods slightly faster following Lars' turn, but it does not clearly mark turn completion, and thereby does not clearly confirm shared understanding. This is evidently an issue for Lars also, as he continues gazing at Bengt throughout the 0.5 second pause (09), before finally initiating a continuation of his telling (10). That is, it has previously been found that when one interactant gazes at another without speaking they are taking the hearer role, and thereby increasing the relevance of a co-participant to speak (Kendon, 1977). By gazing at Bengt, Lars appears to monitor his behaviour for signals of lack of understanding. As Bengt neither confirms nor disconfirms shared understanding, Lars decides to continue on his telling, and thereby assumes that shared understanding is achieved for current purposes.

This example must be understood in terms of the strong sequential constraints for Lars to continue beyond line 08, as his story is not yet complete. Thus anything but displayed hearership from Bengt is not to be expected. However, a more clearly differentiated head-nod at TCU completion might have led to a quicker turn transition from Lars.

5.4.3 Summary

This section has shown that when the development of an anticipatory nod does not proceed as formulated in Table 5.A, this has consequences for how the interaction proceeds. Thus, this demonstrates how interactants are sensitive to the use and extension of anticipatory nodding.

The first example (5.4) showed that stopping a head-nod prior to TCU completion is indicative of a failure in shared understanding, which then calls for modification and repair in a next turn. The following two examples (5.5 and 5.3b) showed that a continued head-nod beyond TCU completion fails to confirm shared understanding, and is used and oriented to as maintaining the display of shared understanding as still interactionally relevant; at the end of a telling (5.5), and between two elements of a telling, where the next element depends on the shared understanding of a previous element (5.3b). Example 5.3b is a boundary case, showing that shared understanding can still be assumed even though it has not clearly been confirmed. The fact that the speaker is in the middle of telling in this example (i.e. strong sequential constraints), might account for how the relevant shared understanding is not addressed any further. Also, this was done as there was no *dis*confirmation of shared understanding.

These examples further support the claim that nodding is only used for as long as no next turn has been initiated.

5.5 Deviant examples

The analysis in sections 5.3 and 5.4 demonstrates the relevance of co-extending an anticipatory nod with the completion of a TCU, for achieving shared understanding. This final analysis section presents instances that deviate from this claim. Two examples will be presented, and in both the anticipatory nodding stops prior to TCU completion. But unlike example 5.4, this does not display any problem in maintaining/achieving shared understanding. However, these deviant features are accounted for, in a way that supports the general argument. In the first of these examples, 5.6, the hearer uses

other means than nodding to display anticipation and understanding. In the second example, 5.7, the lack of head-nods is oriented to as signalling disengagement from current talk.

5.5.1 Alternative means of anticipating shared understanding

Sigurd and Lars are both fans of progressive rock (prog-rock), or art-rock which is the term Sigurd uses here (these are synonymous, which will be relevant for part of the analysis below). In the excerpt below (5.6) Sigurd is explaining how he got interested in prog-rock in the first place. Apparently this happened as he got tired of listening to contemporary metal music, and started listening to Black Sabbath and Led Zeppelin (see lines 01-02). Prior to this example Lars has misheard Sigurd as complaining about the qualities of contemporary prog-rock, when in fact he has been complaining about contemporary metal music. This excerpt follows an extended sequence where Sigurd is trying to resolve this problem in understanding. Lars' displayed alignment (nodding) in parallel with Sigurd's talk in lines 02-04 is therefore relevant to show that shared understanding is achieved.

```
(5.6) KTH-NO, SL, 05:38 "kunstrock"
01   S:            og   da   liksom `GJENnom (eh)  black sabbath og
                   AND  THEN LIKE    THROUGH       name             AND
                   and then you know via   (uh)    Black Sabbath and

                   led `ZEPpe´lin ikke sant,=
                   name           NOT   TRUE
                   Led Zeppelin   right

02   S:      1->   =så  kom  jeg liksom `SAKte  og  gradvis  over ¯PÅ: (.)
                    part CAME I   LIKE   SLOWLY AND GRADUALLY OVER  ON
                    I (slowly) got       more and more into (.)

             2*->  =°hh  den litt   mer [(.)`S]PENnen´de `KUNSTrock´en `da;
                         THE LITTLE MORE      EXCITING      ART-ROCKdet   THEN
                    °hh the (slightly) more (.) interesting art-rock

03   L:                                       [b::  ]

04           3->   (-)

05   S:            °ptkhh og da::::vm:: (--) ↑´ja(m) da   er det jo   et helt
                          AND THEN           YES    THEN IS IT part A  WHOLE
                   °ptkhh and then          (--) yeah(m) then there is a whole
```

```
                   ´H:A:V  (.)  å   ta    av,
                   OCEAN        TO TAKE OF
                   lot (.)  to choose among
```

Lars starts nodding in line 02 during Sigurd's prolonged vowel in *på*/"on" and continues during the pause that follows; i.e. it follows a mid-TCU pause comparable to those described above. Then Lars stops nodding for a while during Sigurd's TCU continuation, before he reinitiates the nodding closer to TCU completion, and then marks turn completion with an intensified head-nod. These details are presented in transcript 5.6a. Two observations in particular will be pursued for further analysis: (i) Lars contexualises his lack of nodding as collaborative, and (ii) these activities seem contingent on Sigurd's gaze.

```
(5.6a) KTH-NO, SL, 05:38 "kunstrock" HEAD-NOD AND GAZE ANNOTATION

STEP                      1                  2              *
02-03  Gz(S)   _____, ,D       , x_____
02-03  HN(L)   ──────────────────────//^^^^^^^^^^^/
02-03  S:              gradvis   over ¯PÅ: (.) °hh  den litt mer [(.)`S]PENnen´de
                       more and more into (.)  °hh the more         (.) interesting

03     L:                                                           [b::  ]

STEP                                         3
02-04  Gz(S)   _____, ,D
02-04  HN(L)   /^^^^^^^^^^^^^^^^/^v^v^v^v^v^v
02-04  S:             `KUNSTrock´en `da; (-)
                       art-rock      (then)
```

In line 03 Lars produces a prolonged "b::", which is a visible (tight lip closure) and audible (voiced bilabial plosive) effort at producing talk. In this case it displays a collaborative effort. A possible candidate for what "b::-" may project is "prog-rock". The voicedness of the bilabial closure might at first seem to counteract such a possibility, since the bilabial in "prog" is expected to be phonetically voiceless. One could argue instead that although Lars initially might have projected "prog-rock", by voicing this closure he is now primarily displaying willingness to participate, whatever the projected word may be. It seems like Sigurd for a brief moment (the 0.1 second pause between *mer* and *spennende*) awaits Lars' verbal contribution. However, Sigurd proceeds, and

instead of projecting further efforts to speak, Lars starts nodding again as Sigurd's TCU nears completion (during *kunstrocken*/"(the) art rock"). Thus, Lars continuously contextualises his actions as collaborative. During the absent nodding he shows that this is not a lack of understanding/alignment by projecting speech material instead.

Both Lars' speech projection and his nodding seem highly contingent on Sigurd's use of gaze, as (i) Lars stops nodding right after the point where Sigurd gazes away from him, and (ii) Lars produces the "b::" right after Sigurd gazes back him. In other words, this example suggests that absence of gaze might reduce the relevance for a continued head-nod.

5.5.2 Early shift to a next turn

The non-presence of head-nods in co-extension with a TCU completion might be accounted for in other ways, that are also not indicative of trouble in understanding. In example 5.7 the absent head-nod signals early turn transition (speaker change), as a means of returning to more pressing matters than those that are currently being presented in talk. As will be shown in this example, a current speaker may also design his talk to accommodate such a transition.

Bengt and Lars are talking about Lars' guide friend in Athens (this interaction is prior to that of examples 5.2 and 5.3). After initially having confirmed that the friend is from Athens (*mm* in 02), Lars modifies his confirmation in line 04-05. Here he elaborates on the information that might (dis)qualify the friend as knowing Athens well enough: The friend does not originally come from Athens (also, Lars' answer in 04 is not accurate in terms of Bengt's use of *hjemby*/"home town" in 01). The friend's origin is clearly not the main objective of Bengt's action initiated in 01, and in 07 he shows that he would rather like to hear more about what they saw and did. Thus in sequential terms, Bengt deletes the relevance of whether the friend was from Athens or not, and the early stop/non-existence of anticipatory nodding plays an important part in shaping this projection (see details in transcript 5.7a).

```
(5.7) KTH-NO, BL, 04:10/382 "Athen"

01   B:              °h men det er `HENnes `HJEMby    så    hun:=
                       BUT IT  IS  HER      HOME-TOWN part SHE
                     °h but it's    her      hometown  so she

02   L:              =mm.=
                      mm

03   B:              =[skulle `VI´se litt   (og)-]
                      SHOULD   SHOW  LITTLE (AND)
                     was going to show a little and

04   L:               [´NEI:   hun?         ´nei ] hun var fra  thessalo`NI´ki,
                        NO      SHE           NO    SHE WAS FROM name
                         no she               no she was from Thessaloniki

05   L:    1->   men eh: dnh° hun hadde ˇBODD: en del  i (.)
                 BUT         SHE HAD    LIVED A  PART IN
                 but uh:      she had lived some time in (.)

           2*->   [ath]en <<all >også(m).>
                   name         ALSO
                  Athens too

06   B:    2*->   [mm-]
                   mm

07   B:    3*->   var `DET  ´kult da?
                  WAS  THAT  COOL THEN
                  was that fun




(5.7a) KTH-NO, BL, 04:10/382 "Athen" HEAD-NOD AND GAZE ANNOTATION


STEP              1                    2    *
05-06 Gz(L)       _____
05-06 Gz(B)       _____+ _____
05-06 HN(B)                                 //^^//
05-06 L:          ˇBODD: en del  i (.) [ath]en <<all >også(m).>
                  lived some time in (.) Athens too

06   B:                                   [mm-]
```

Bengt's response appears quite terse in this context, since he does not continuously display attention to the development of the TCU-completion, like in the above examples. This does not create any trouble regarding shared understanding in this example. Instead, Lars seems to acccommodate Bengt in projecting a turn-transition. Lars does so by increasing his speech rate during the last word of the turn (*også*/"also"). A decreased rather than increased speech rate is expected at the end of utterances/turns (Abercrombie, 1964), however it has previously been shown by Local

and Walker (2004) that a turn-final increase in speech rate can be used as a resource for projecting a next turn (what is referred to as 'rush-through' elsewhere in the CA literature; e.g. Schegloff, 1996b). These studies focussed on projections by a single speaker, whereas example 5.7 suggests that increased speech rate can also be used collaboratively, to leave way for a co-participant's next turn.

Furthermore, Lars' closing of lips upon completion and a following (slight) head-nod appears to form further turn-yielding cues. A representation of Lars' TCU completion is given in Figure 5.A below, including durations of syllables (in ms). We see that the last two syllables *også*/"also" are approximately 1/3 to 1/2 the duration of the previous syllables (the last vowel of *også* is realised with a single glottalised pulse). It is also noticeably quieter than the preceding context.

SYLL:	en	del	i	(.)	At-		hen	og	så
DUR(ms):	182	145	142		130		189	40	96
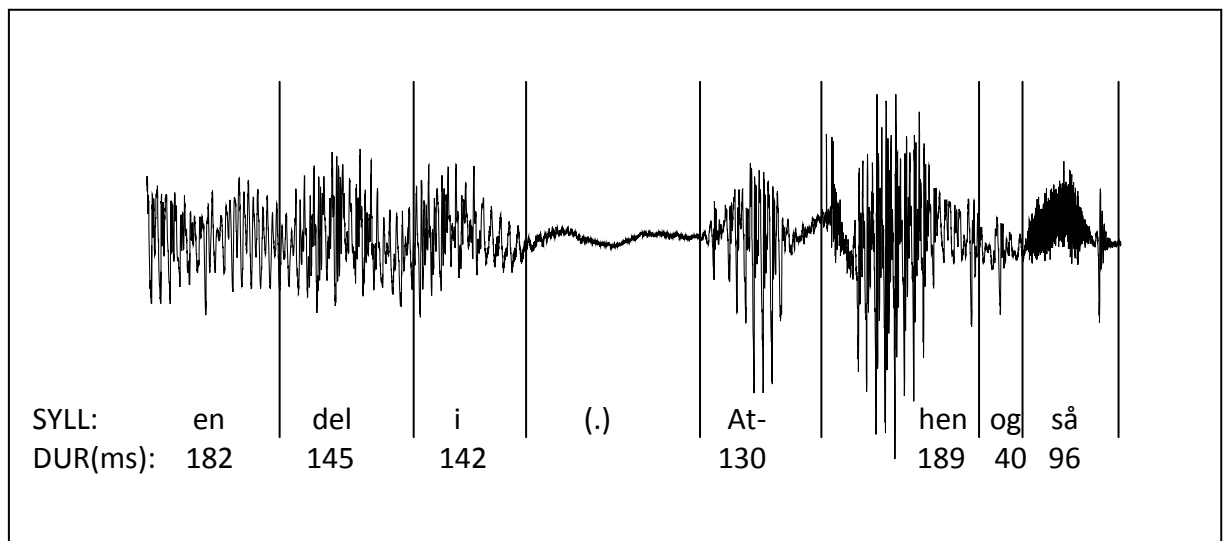
Figure 5.A. Waveform representation of Lars' *en del i (.) Athen også*/"some time in (.) Athens too". Separated into syllables (SYLL). Durations (DUR) given in milliseconds (ms).

What makes this example different from the previous examples is that the interactants are negotiating shift from one action to a next one, while handling a multi-unit turn. Bengt displays this projected shift of speaker and turn by not nodding, i.e. the display of shared understanding is abandoned in favour of Bengt's projected shift.

141

5.5.3 Summary

This section has presented examples that deviate from the norm presented and demonstrated in the previous sections. Instead of arguing against the initial claims regarding the relevance of anticipatory nodding, it supports the analysis in enriched ways. The first example, 5.6, showed how a hearer may use alternative means to display anticipation of understanding, in this case by projecting a verbal collaboration instead of a non-verbal one. Thus the essential factor in displaying shared understanding is the presence of a collaborative action, not whether or not there is nodding. However, nodding seems to be the default resource for maintaining shared understanding during the production of a turn.

Example 5.7 showed that the absence of anticipatory nodding may also be oriented to as projective of a next turn, and thus disengaging with the current turn, and treating the display of shared understanding as no longer relevant. This shows how a hearers' alignment with a current turn and a projection of a next turn can overlap, and that the interactants manipulate resources (in this case the absence of anticipatory nodding) to signal this.

5.6 Summary and discussion

The objective of this chapter was to demonstrate one way in which the speaker and hearer depend on each other's actions in the production of a turn, and that this joint achievement has consequences for what follows. The practice in focus has been one where a speaker makes the hearer's displayed understanding relevant in the middle of a TCU, and the hearer uses head-nods to display anticipatory understanding of the turn while it is still in progress. The analysis has shown that the hearer's anticipatory nodding does not only facilitate the speaker's further turn production, its exact extension with the speaker's TCU is crucial for the maintenance and achievement of shared understanding. If the nodding stops prior to, or extends in an undifferentiated manner beyond the TCU completion, this has observable consequences for the subsequent talk.

A set of deviant examples contributed further evidence by showing how an absence of co-extensive head-nods needs to be accounted for in certain ways to avoid displaying trouble in shared understanding.

One obvious point to make based about these findings, is how interdependent social achievements are with structural constraints in interaction. Here the social achievement of shared understanding is at the same time an achievement of turn completion, because the TCU is constituted by a speaker's verbal-syntactic completion of an action that promoted the display of shared understanding. It is for this reason that nods must co-extend with the TCU to constitute understanding. But also speakers' and hearers' actions are interdependent, or contingent, in this process. It is based on their joint orientations during the production of the TCU that turn completion, and shared understanding, can be achieved. This means that not only are traditional categories like TCU a unit constitutive of an *action* (Sacks, et al., 1974), but also a unit built on continuous *inter*action.

Normally one expects to find hearer's display of alignment/understanding at the end of a TCU. This study shows that a TCU can be used more flexibly than such, to secure shared understanding during a turn, and prior to turn completion. This does indeed show one way in which interactants' 'hidden work' (cf. Heritage, 2007) establishes shared understanding, i.e. how shared understanding is not simply assumed, but based on detailed, co-ordinated interactant work. This is not meant to imply that speakers never simply assume shared understanding, but there might be further constraints and resources used to visibly/audibly secure the understanding, that have not been considered or studied yet.

In relation to the study by Stivers (2008) it is interesting to note that the head-nods in this study also can be conceptualised as displaying access to a speaker's stance, also for other sequential environments than the end of story-tellings. Furthermore, it is not only the presence of nodding that seems relevant in displaying access to speaker's stance, the timing and co-extension of nods with the ongoing talk is crucial for maintaining this displayed access.

As an interactional resource, nodding might be rather well-fitted to display understanding during turn production, as it is visual and does not interfer with the speech signal. I would expect that the particular use of head-nods reported on here might be found across cultures and languages, as a common resource for aligning speech production with shared understanding. Although one might argue that head-nods are non-linguistic, it is not straightforward to distinguish them from verbal alignment tokens like "yes": They may both constitute an agreement, or an alignment with ongoing talk. In terms of social action, head-nods are one resource with which interactants contextualise a verbal turn production as achieving certain things, which in this thesis is conceptualised as part of language.

A further point worth making is that the halts, prolongation and reversed co-articulations observed (i.e. the production *femtenåringen* in example 5.3), are by no means 'speech errors', but interactional resources, which interactants pay careful attention to in the emerging talk. This, like Goodwin (1981) stresses, demonstrates the importance of understanding speech as integral to the interactional process in which it is embedded.

There may be several issues which this study does not address, and which would benefit further exploration. One such issue is how precise the co-extension of nodding needs to be with the completion of a TCU to constitute shared understanding. Example 5.4 suggested that trouble is displayed only by stopping to nod a syllable prior to the TCU completion. However, a more precise formulation and testing would have to be done in a further study.

Another issue to pursue further is exactly what constitutes the relevance of a mid-TCU response. It is clear that speakers may create a pause in the middle of a TCU to establish shared understanding, but one might ask whether such pauses are most relevant to two-parted structures (e.g. "when-then"). Most instances reported here were two-parted structured, but the question remains whether a mid-TCU response can be made relevant at any point of an emerging TCU.

It was found that a speaker typically triggered a response with prolongations, halts, particles like *så*, and non-coarticulatory articulations, while maintaining mutual gaze. In

response the hearer produced only minimal alignment tokens (verbal + nod). Is it provided in this design that hearers will not contribute explicitly, as they are not for example searching for a word, or in other ways asking for explicit assistance? What would an invitation of explicit assistance look like? This last question will be the topic of the next chapter, where we will see how gesture holds can be used as part of such an explicit request. As in the current chapter, chapter 6 will focus on how the timing of non-verbal resources with speech is of central relevance for the achievement of shared understanding.

CHAPTER 6

GESTURE HOLDS AND RESOLVING SHARED UNDERSTANDING

This chapter continues to investigate speakers' and hearers' parallel actions, by focussing on manual gestures (which will be referred to simply as 'gestures') as a resource for joint achievements in talk. The particular phenomenon studied here is gesture hold, i.e. where a speaker holds their speech-accompanying gesture beyond the verbal completion of their turn, and into a next turn produced by a co-participant. In general, by extending their gesture in this way interactants (i.e. gesturers) display an orientation to the development of a projected action, or understanding. In this way the gesturer displays qualities of being speaker (i.e. gesturing), while being hearer (expecting response from co-participant)

In many cases this use of gesture displays that there is an issue with understanding, and that this issue needs to be explicitly brought forward to the surface of interaction, i.e. the co-participant's assistance is needed to resolve the issue. Such cases will be the focus of this study, by providing detailed analyses of how interactants orient to the initiation, maintenance and extension of such action with the use of gesture hold. In particular the study will show how the precise timing of gesture holds is crucial to resolving shared understanding. Like chapter 5, this chapter addresses the interactional achievement of shared understanding. But where the previous chapter addressed the maintenance of implicit shared understanding, this chapter focuses on how shared understanding is explicitly brought to the surface in the interaction, and how gesture holds form a particular resource in this regard.

Example 6.1 is given below as an illustration, and as a starting point for the study. Lars talks about people's attitudes towards musical scratching: In general they do not regard scratching as music (scratching is using turntable as a musical instrument; see also entry 'scratching' in Appendix C). Of main interest is Tor's turn (lines 03-08), where he checks

whether his understanding is correct with *ja å scratche*/"(yes) to scratch", accompanied by a gesture and gesture hold. Tor holds his gesture until Lars has confirmed Tor's candidate understanding.

Tor's gesture/hold is illustrated in still-shots (a-d; the zig-zag lines indicate with what element in the verbal construction each separate still-shot is aligned), and transcribed above the verbal components. The annotation conventions for gestures are explained in chapter 3 (section 3.4.3). In short, '^^x' represents main part of gesture (stroke), where 'x' represent the peak of the gesture; '....' represents early phases/preparation of a gesture if preceding '^^x', and release of gesture if following '^^x'. And finally, '----' represents gesture hold.

```
(6.1) KTH-NO, TL, 14:07 "scratche"

01     L:              og `FORTsatt så   er °hhhh er det `VELdig vanskelig å:
                       AND STILL    part IS       IS IT   VERY   DIFFICULT TO
                       and still it's       °hhhh it's very difficult to

                       å `OVertale folk << all > om    at   det ja  men det>
                       TO CONVINCE PEOPLE           ABOUT THAT IT  YES BUT IT
                       to convince people           that            "yes but

                       `ER ↑↑jo mu`sikk.
                        IS part MUSIC
                        it IS music"

02                     (1.0)
```
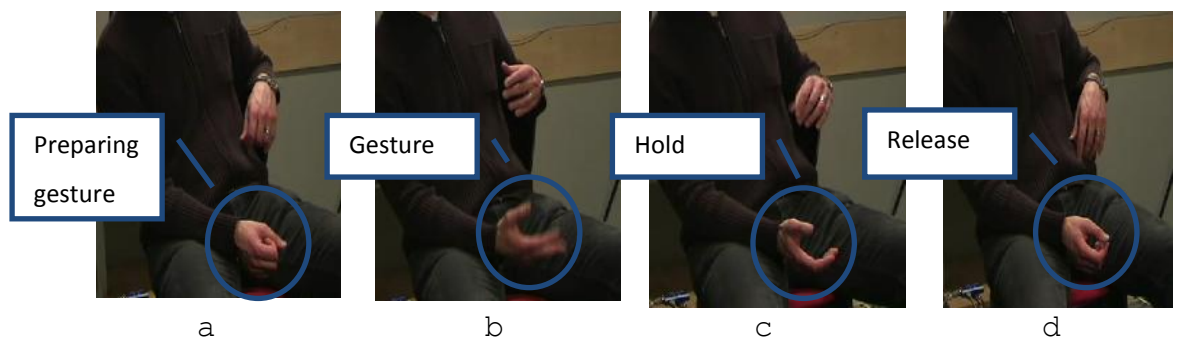


```
03-08  MG(T)          ..............^^^^^^^^^^x--------------....(..)
03-07  T:             °pth[h    ] ´ja(ˀ)å [`SCRAT]´che,=        (.) mm=
                              YES   TO   SCRATCH
                      °pthh       yes to scratch              (.) mm

04-08  L:                    [°pth]        [forˀ  ]    =(n)ˇja,     =fordi
                                            BECAUSE         YES        BECAUSE
                             °pth          becau-          yes        because
```

In 01 Lars assesses scratching as music, but that it is hard to convince others to think the same. In general such claims about the status of a referent, makes relevant some sort of agreement from co-participant (Pomerantz, 1984). In response Tor indicates some trouble in providing such agreement (there is a 1 second gap in 02), and with *ja å scratche* Tor requests Lars' confirmation that his candidate understanding is correct.

Notice that Tor, after having reached the peak of his gesture with the first syllable of *scratche*/"scratch" ('^^^x'; figure b-c), holds his gesture ('---'; figure c) until Lars has produced the confirmation *ja*/"yes". Then Tor releases his gesture ('...', figure d) while he produces a verbal validation *mm*. Thus Tor's gesture hold is co-extensive with a minimal sequence in which Tor makes relevant, and receives, a confirmation from Lars. Also Lars' actions display an orientation to this as being a minimal sequence: Immediately following Tor's verbal validation *mm*, Lars reinitiates the *fordi*/"because" that was first initiated in overlap with Tor's *scratche*.

Tor's gesture hold seems to both project and 'await' a confirmation from Lars, before accepting the confirmation with a gesture release. This chapter will show that this is the case for a range of instances, and that the timing-relations between gesture hold and verbal elements follow a general shape that is recognisable to participants as projecting and displaying shared understanding. The general shape is formalised in Table 6.A. As in chapter 5, this sequence will be the point of reference for the case-by-case analysis that follows.

Table 6.A. Formalisation of the sequence of events which lead to the achievement of shared understanding, separated in three steps (columns) and between speakers (rows).

Speaker	Step 1	Step 2	Step 3
A	*Speaker*: Brings an issue regarding understanding to the surface of interaction (e.g. understanding check): Using verbal resources accompanied by gesture	*Speaker/hearer*: Orients to speaker B's contribution, while holding gesture	*Speaker*: Displaying achievement of shared understanding. Releasing gesture followed by verbal response
B	*Hearer*	*Hearer*: Produces contribution to shared understanding (e.g. a confirmation)	*Hearer*

Before attending in detail to the use of gesture hold in resolving shared understanding, a background to the study on interactional gestures will be provided (section 6.1), followed by a description and an overview of the range of interactional uses gesture holds are involved in (section 6.2). As will be shown here, gesture hold is often associated with interactional problem-solving. This forms the basis for sections 6.3, where a set of instances will further demonstrate the process in which gesture hold is used to bring an explicit understanding to the surface of interaction, and then display its resolution. With a set of negative examples, section 6.4 further defines this action by, (i) showing that gesture holds are *not* used when co-participant's assistance is not projected or relevant, and (ii) showing that the interactional relevance of gesture hold requires the presence of mutual gaze. Section 6.5 provides a set of examples showing that the timing of releasing gesture hold might not always be as schematised above, but that this occurs in orientation to particular interactional developments.

6.1 Background: Interactional gestures

One important conclusion to be drawn from previous studies on speech-accompanying gesture (e.g. McNeill, 1992, 2005; Kita, 1996; Goldin-Meadow, 2003; Kendon, 2004) is that gesture and speech are co-ordinated in certain ways and compensate each other in order to express meaning. However, most of the available literature focuses primarily on individual processes (i.e. utterance production) rather than interactional processes that gesture production is part of (cf. chapter 2, section 2.3). Although it is a common observation that gestures regularly elaborate, refine and/or highlight verbal elements in talk (Kendon, 2004), a much less common observation is that also hearers (i.e. interactactants who are not currently speaking) gesture, and co-express elements in a co-participant's talk. This study will emphasise the relevance temporal and interactional processes in studying how and when gestures communicate.

This section introduces some of the (few) studies that demonstrate the interactional role of gestures. First, 6.1.1 will give a background on how gestures have been conceptualised and explored as interactional in previous research. Further, 6.1.2 will review previous findings that account for how gestures may be used in turn-taking, i.e. to project a next turn. This will include some observations on gesture holds. Few previous research looks into the precise timing of gesture with speech in defining interactional processes, and achievements, and this will be one main contribution of the current study.

6.1.1 What are interactional gestures?

An issue continually documented in studies on speech and gesture has been to what extent gestures are communicative, i.e. whether or not they make a difference for the interactional management, and for a listener's understanding. A range of experimental and observational studies have now provided a strong basis for claiming that gestures do communicate (see especially Bavelas, 1994; Kendon, 1994). However, much work

remains to show how and when gestures communicate, i.e. how gestures become relevant for speakers and hearers while they interact (Kendon, 1994, 2004).

Bavelas et al. (1992) were among the first to systematically explore the potential that lies in exploring interactional, or what they called *interactive*, gestures. According to Bavelas (1994), the interactive gestures count for about 15% of the gestures we use. Bavelas et al. (1992) distinguished between interactive and *topical* gestures, where interactive gestures were defined as having no content function, but rather involving the addressee in one way or the other. Topical gestures on the other hand seemed to capture representational meaning. They found that the amount of such gestures depended on (i) whether there was an addressee present, and (ii) whether the speaker and addressee could see each other or not. Based on these results, Bavelas et al. (1992) and Bavelas (1994) claim that interactive gestures have distinct communicative functions.

Bavelas (1994) lists a range of functions interactional gestures can have, including sharing or acknowledging information, seeking agreement and giving the turn away. But a short-coming of Bavelas' description is that she does not pay particular attention to what it is in the gesture that serves or creates an interactional function, and thus the distinction between interactional and topical gestures is not entirely clear. It is not explicitly explored for example, whether so-called topical (representational) gestures can also be interactive. That is, do topical gestures take particular shapes, different from interactional ones? Or can topical/representational gestures be used and manipulated in such a way that they 'become' interactive? I would argue that a further insight into the interactiveness of gestures must come from a more detailed analysis on how gestures are used and become relevant to speakers and their co-participants. Some studies which address this concern will be reviewed next.

6.1.2 Gestures and gesture hold in the management of turns at talk

There are studies that explore the use of gestures (and other non-verbal behaviour) by focussing in more detail on the interactional process itself, but these are still rare

compared to studies focussing on verbal conduct. Some such studies show how gestures may be used in projecting turn transition (Streeck & Hartge, 1992; Mondada, 2007), and during word searches (Goodwin & Goodwin, 1986; Hayashi, 2003b; Streeck, 2009). Streeck and Hartge (1992) observed that gestures may be used as an 'entry', or preface, to an upcoming turn of talk. They studied two different gestural moves in the language of Ilokano (Philippines), one involving a palm up manual gesture, the other open mouth ('[a] face'). Both gesture types were associated with the projection of turn transition, but the palm up gesture was found to be more specific to tellings than the [a] face, for example where a participant would signal that he/she had more to tell on a story.

Previous accounts of gesture holds have mainly focussed on their utterance-internal use. McNeill (2005) for example, describes gesture hold in terms of how a gesture is held prior to or following a co-expressive verbal element, in order to ensure temporal co-ordination between the two. Also, Liddell (2003) shows how signers of ASL (American Sign Language) sometimes use *buoys* in signing: In the case of one sub-type of buoys, pointers, this involves a point on the non-dominant hand to direct attention to an entity while the dominant had is used to sign other information (Liddell, 2003, pp. 250-260). Pointers such as these have a gestural element in that the location and extent of the point is determined by context, rather than being a grammatical feature.

The primary motivation behind this study is the role gesture plays in continuing an action beyond the verbal completion of a turn and into a next turn. Thus, we are going to focus on a particular usage of gesture hold, which can be set aside from that of many previous accounts of this concept. Also, the current focus can be set aside from previous studies on the use of gesture to build cohesion multi-unit turns, or utterances (e.g. McNeill, 2005; Enfied 2009), since these do not explore gesture holds, or the immediate linking between one turn and a (co-participant's) next turn.

There are a few accounts in previous studies of gesture holds as used beyond the boundary of a turn/utterance. First, Kendon (1995) studied gestures as a question-marking feature (in Southern Italian). As part of his study, he reported on some instances where the gesture continued well beyond the point the speaker's turn

finishes, and argued that this use of gesture served to make clear that what has just been said is a question which requires an answer. Another interactional description of gesture hold across turns is found in Mondada (2007). She focussed on the use of pointing in projecting a next turn, providing detailed analysis of how the projective nature of the pointing was negotiated in real time. As part of her study she investigated an instance where the pointing gestures persisted after the projecting speaker's turn completion. This pointing gesture was accompanying a question, and its producer would hold the gesture until the end of the answer, and "stopping just before her [the pointer's] acknowledgment" (p. 216). Thus, Mondada's example is an interesting parallel to example 6.1 presented above, where the gesturing speaker (Tor) also released the gesture just before his acknowledgment.

Unlike the current study, Mondada (2007) and Kendon (1995) do not go further to explore the timing relations between the gesture hold and other elements in the sequence. Although Mondada (2007) argues that the hold (i.e. in the example reported on above) orients to the adjacency pair as the relevant sequential unit, she does not address the potential generalisability of gesture hold in terms of the particular actions that interactants perform.

As we saw in chapter 2, one set of findings argues for investigating a range of semiotic resources rather than particular ones, when addressing the management of talk-in-interaction (e.g. Goodwin & Goodwin, 1986; Goodwin, 2000). One finding that is particularly interesting in context of this study is how gaze may affect the interactional relevance of gestures. Investigating how speakers perform word searches, Goodwin & Goodwin (1986) found that they regularly look away from their recipient during the word search. The speaker may involve gestures in the word search, but the recipient will only display an understanding of this gesture when the speaker turns gaze at him/her. Thus, gaze seems to be a powerful way of triggering co-participation, which may or may not involve the use of gesture. This phenomenon has later been studied and described by Hayashi (2003a, 2003b) and Streeck (2009). Streeck (2009) also demonstrates that gazing at own gesturing hands is an interactional resource for attracting a co-participant's attention to, and potential contribution to, talk in progress. In other words,

what is interactional about gestures is not only defined by the gestures themselves, but by the way they are employed along with other resources.

6.1.3 Summary

Research on the interactional use of gesture is relevant as we still know very little about how to describe such gestures, in terms of what they are and how they are relevant for speakers and their co-participants. It is not enough to show *that* gestures have interactional value, rather we need to explore *how* they form interactional value. Although sporadically reported on, gesture holds across turns of talk have not been studied systematically as an interactional phenomenon. Such use of gesture holds are a particularly interesting resource because they offer one way in which interactants can in a sense 'continue' their turn into a co-participant's turn, and as such display orientation to the relevance of the next turn in relation to the previous one.

This study follows a handful of previous studies, which look especially at speech-accompanying gestures, in relation to the achievement of interactional goals. This study will contribute to the extant literature on gesture and social interaction, in demonstrating one way in which the interactional role of gestures is determined by their fine, temporally unfolding, co-ordination with speech.

6.2 An overview of gesture hold used across turn-boundaries

This section describes the processes involved in finding the set of comparable instances of gesture hold used in this study, and will work as an introduction to the kind of actions gesture holds take part in. The procedures for collecting and categorising instances will be described further in 6.2.1. Examples of action categories and a general overview of the distribution of gesture holds across action types will be presented in 6.2.2 and 6.2.3, respectively.

6.2.1 Procedures

In order to support a rich account of the phenomenon of gesture hold across turn boundaries, all instances of gesture holds were collected and studied in the 80 minutes of free conversation in the Norwegian data. This was done manually, using ELAN (see chapter 3, section 3.4.1). There were no restrictions as to what type of gestures would count as a gesture hold, e.g. iconic, indexical, metaphoric (McNeill, 1992). This was done because this study, and the phenomenon at hand, deals more directly with the co-ordination of gesture with speech, than with the conceptual relation between gesture's form and verbal-propositional content. I did not expect that gesture type itself would affect the way in which gesture holds were co-ordinated with speech.

For purposes of comparison, instances where gesture was used, but *not* held into co-participant's turn were also collected. The definition of a gesture hold was that it would be held beyond a speaker's verbal conduct and maintained during a co-participant's talk. This could be in the middle or at the end of a TCU, and the definition of a turn-transition was that the co-participant's next talk would form a more-than-minimal contribution to the interaction. That is, they would form a next speaker action, in the form of a Second Pair Part, or in other ways add substantially to the ongoing talk (e.g. comment on the speaker's talk). This excluded gesture holds in overlap with short responses such as "mhm" and head-nods for example (as we have seen above, these

are more readily categorised as a hearer action, rather than a speaker action). To address the distribution of gesture holds, all instances of turn-transitions according to the above definitions were labelled.

The instances were analysed separately (techniques and conventions for conducting interactional and gestural analyses are reported in chapter 3). As part of this process, some action categories emerged and were further developed for purposes of providing a general overview of the phenomenon. The action categories used were as follows:

- Understanding check
- Understanding request
- Clarification request
- Word search
- Seeking agreement
- Incidental incomings
- Other/miscellaneous

These categories helped to set the boundaries for which instances would form the main basis for the current study. Also, they were used to address the proportion of gesture hold according to actions category. This will be described further in 6.2.3. Further descriptions of the action categories are given in appendix D.

6.2.2 Examples of gesture holds according to action categories

This subsection illustrates the categories *understanding request*, *seeking agreement*, *word search* and *incidental incoming* each with an example. In general, gesture holds display that some action is not yet complete, and by holding a gesture while a co-participant talks its producer displays both aspects of speakership *and* hearership to the incoming talk. In some instances gesture hold comes about as one participant projects a specific contribution from another, in other instances the gesture hold is responsive to incoming talk from co-participant. In most instances gesture holds appear to monitor specifically toward a projected outcome, or shared understanding. The exceptions to

these generalities are instances where the incoming talk is not really fitted with the projected meaning, described as *incidental incomings* below.

Whereas example 6.1 presented in the introduction was an example of *understanding check*, example 6.2 presents an instance of what was categorised as *understanding request*. Here Anne seeks a confirmation from Oscar of whether his daughter knew her daughter from high school (line 02). Oscar responds with a multi-unit turn (05), in parallel with which Anne holds her gesture until the point when Oscar disconfirms Anne's proposal. The similarities between example 6.1 and 6.2 is that the speaker specifically seeks a confirmation from the co-participant. Unlike 6.1 though Anne does not check her own understanding as such, she proposes something (the potential connection between their two daughters) which requires (i.e. requests) Oscar's assistance and knowledge to complete.

At this stage I will mainly pay attention to the presence/non-presence of gesture hold. The relevant turns where gesture hold is initiated are highlighted with an arrow, and the gestural transcriptions will only include the start/end of a gesture unit ('//'), and the presence of gesture hold ('---').

```
(6.2) KTH-NO, AO, 16:03 "Kungsholmens musikklasser"

01    A:            ´JEg må   høre me:d; (.) n `D:Øt:      med  den eldste
                     I   MUST HEAR WITH       DAUGHT(ERS) WITH THE OLDEST
                    I have to ask my daught(ers)- my oldest

      MG(A)                             //
                    ´DAtteren   min  som `GIKK <<p > ↑på> °h (xxx xxx)
                    DAUGHTERdet MINE WHO  WENT        ON
                    daughter who went to °h (xxx xxx)

      MG(A)                                            /HOLD--
02          ->      altså [i ´K]Ungsholmens mu`S[IK´kla]sser.
                    SO    IN name-gen        MUSIC-CLASSES
                    you know Kungsholmens music course

03    O:                    [(ˀ)]                [jahaˀ]
                                                  YES
                                                 oh right

      MG(A)         -------------
04    O:            << f >´JAha.>
                          YES
                         oh right
```

```
        MG(A)            ---------------------------------------/
05      O:      ->   okay; (h°) (s) de:t ^DEt  er ikke sikkert at  de:; (.)
                     OKAY           IT  IT  IS  NOT CERTAIN THAT THEY
                     okay (°h) it it's not certain that they

        MG(A)            //
06      A:           nei.
                     NO
                     no
```

Verbally Anne's turn in 01-02 is not clearly formulated as a request for a contribution from Oscar, although her use of *altså*/"you know" might enhance this potential. Thus, Anne's use of gesture hold from 02-05 seems to play an important part in Oscar's orientation to this as a request for confirmation. For example, Anne continues her gesture hold beyond Oscar's first verbal response *jaha*/"yes" and *okay*, showing that these did not really resolve her understanding request. It is not until Oscar has clearly disconfirmed Anne's proposal that their two daughters knew each other: Anne releases her gesture hold after *det er ikke sikkert at de*/"it's not certain they" (05).

Seeking agreement. Example 6.3 shows that gesture hold can also appeal to co-participant's agreement. In 01 Sigurd provides a negative assessment of the band "Meshuggah" (referred to by *det*/"it"). Sigurd accompanies this turn by a gesture that is held during the gap that follows (02), and during Lars' response (03), where it is made clear that Lars disagrees with him, at which point Sigurd releases his gesture. In correspondence with 6.1 and 6.2, by holding his gesture Sigurd appears to both project and monitor a potential agreement from Lars.

```
(6.3) KTH-NO, SL, 16:09 "Meshuggah"

01      MG(S)                //                          /HOLD--
01      S:      ->   det  synes `JEG er litt    kjedelig da—
                     THAT THINK  I   IS A-LITTLE BORING   THEN
                     I think that's a bit boring

02      MG(S)            ---
02                       (-)

03      MG(S)            --------------------/
03      L:           ´JEG eh `LI:ker meshuggah. H°
                      I     LIKE    name
                      I uh like Meshuggah H°

04      MG(S)            //
04      S:           [ja okay, ]
                     YES OKAY
```

158

```
                              yes okay

05      L:            [jeg synes] de  er `SKITbra—
                       I   THINK  THEY ARE SHIT-GOOD
                       I think they are really good
```

Word search. In example 6.4, Sigurd has indicated trouble in finding a word in 04/06, which appears to be the name of a person in the band "Panzerpappa" (*det er han eh*/"it's him uh"; *han som driver*/"he who runs" in 04). During the prolonged *det derre::*/"that:" in 06 Sigurd initiates a gesture, which is held after Lars initiates the candidate *Trym Skjevstad* in 07 (specifically the gesture hold starts right after the release of alveolar [t] in *Trym*).

```
(6.4) KTH-NO, SL, 11:14/674 "Panzerpappa"

01      L:            har  du  hørt  om eh `PANzerpappa [forres´ten?]
                      HAVE YOU HEARD OF    name         (BY-THE-WAY)
                      have you heard about uh Panzerpappa by the way

02      S:                                        [°hh        ]°hh
                                                   °hhhh

03      S:            =eh::m:: °tk ´JA h° det  `HAr  ´jeg,
                                   YES    THAT HAVE  I
                       uhm      °tk  yes I have

04      S:            det er `HAn eh::m::: mh° °ptk [h    ]an som driver=
                      THAT IS HE                     HE       WHO RUNS
                      that's the one uhm mh° °ptk the one   who runs

05      L:                                         [(eh) ]
                                                    (uh)

06      MG(S)                    //   /HOLD--
06      S:     ->    =det derre:[: vmb      ]
                      THAT (THERE)
                      that:
07      MG(S)                          (HOLD )-------/   //
07      L:                   [°p trym `SKJ]EV´sta[d, ]
                                        name
                                        Trym Skjevstad

08      S:                                        [d  ]

09                    (.)

10      S:            j ja mh° stemmer(m);
                      YES   CORRESPOND
                      y- yes exactly
```

Notice that Sigurd's gesture hold starts at the same time as he halts speech production (*vmb* in 06), immediately after Lars initiates his incoming 07. In this way Sigurd abandons his turn production *while* maintaining the projection of a word search, and he does so in orientation to Lars' incoming response. Notice also that that Sigurd holds his gesture until near the end of Lars' candidate (07). In this way 6.4 supports the general claim that gesture holds display hearership, while monitoring (and displaying sensitivity to) the co-participant's potential contribution towards a projected outcome. A difference between example 6.4 and examples 6.1-6.3 however, is that the gesture hold is initiated *following* a co-participant's response. In example 6.4, this reflects that Lars' assistance occurs at a point (i.e. the middle of a TCU) when it is not clearly defined whether or when Lars *should* make his contribution. In other words, Sigurd does not clearly project Lars' contribution. However, as Lars does contribute, Sigurd orients to this contribution in much the same way as in the previous examples.

Incidental incoming. Also in example 6.5 the gesture hold is initiated after incoming talk by co-participant. However, unlike example 6.4 the gesture hold is responsive to something more 'incidental' to the current talk, i.e. it does not collaborate on the gesturing speaker's projected meaning but rather directs attention to previous matters.

Lars is talking about his experiences in Athens, where there had recently been some riots that resulted in a high increase of police force in the streets. In 03 Lars makes use of a gesture to illustrate how the police wore shields in the streets. At the same time, in 04, Bengt initiates a news-receipt, *nå*/"now". Clearly, the news-receipt is not about co-projecting Lars' action produced in overlap (*med skjold*/"with shields" in 03), but rather addressing the issue of having police in the streets. In response to Bengt's news-receipt Lars aborts his speech production *med skjo-*/"with shi-", while holding his gesture. Lars then holds his gesture while producing a confirmation *ja*/"yes" directed at Bengt's news-receipt, and following that reinitiates his gesture (indicated by (*) in the transcript).

```
(6.5) KTH-NO, BL, 04:35 "med skjold"

01    L:              °ptk på ´HVErt eneste `HJØr´ne,
                         ON  EVERY SINGLE  CORNER
                      °ptk on each corner

02                    <<all >gatehjørne    så> `VAR de:t— to: (eh) poli`TIer
                         STREET-CORNER part WAS IT   TWO        POLICEpl
                         street corner there were two police-men

03    MG(L)           //        /HOLD----------------/(*)
03                    me:d h° °h[h  ] med  skj    ^JA med  ´SKJOld o:g
                      WITH           WITH (SHIELDS) YES WITH SHIELDS  AND
                      With h° °hh     with shie- yes with shields and

04    B:                           [nå,]
                                    NOW
                                    now
```

This is then another instance where a speaker, while gesturing, displays hearership towards co-participant's talk. But one cannot straightforwardly claim that the gesture hold here is about monitoring co-participant talk for its contribution towards a projected outcome. Example 6.5 is different from 6.1-6.4 in that the co-participant does not contribute to the projected meaning. What Lars shows by holding his gesture though, is that his action is not yet complete, which is further confirmed by the following re-initiation of his gesture and verbal production.

In sum, what we have seen in these instances is that gesture holds in general display incompleteness of some action, and hearership. In some cases gesture holds appear to project a co-participant response (examples 6.1-6.3), in other cases they appear responsive to co-participant talk (examples 6.4-6.5). Finally, in most cases (except example 6.5) gesture holds appear to display the progress of shared understanding of a current project, by monitoring towards it.

6.2.3 Distribution of gesture hold across action categories

There were in total 41 instances of gesture hold as defined in 6.2.1. Figure 6.A shows the distribution of these according to action categories. This distribution shows that there were relatively few word searches involved (7%), compared to understanding

check (19%), understanding request (27%), clarification request (15%) and incidental incomings (15%).
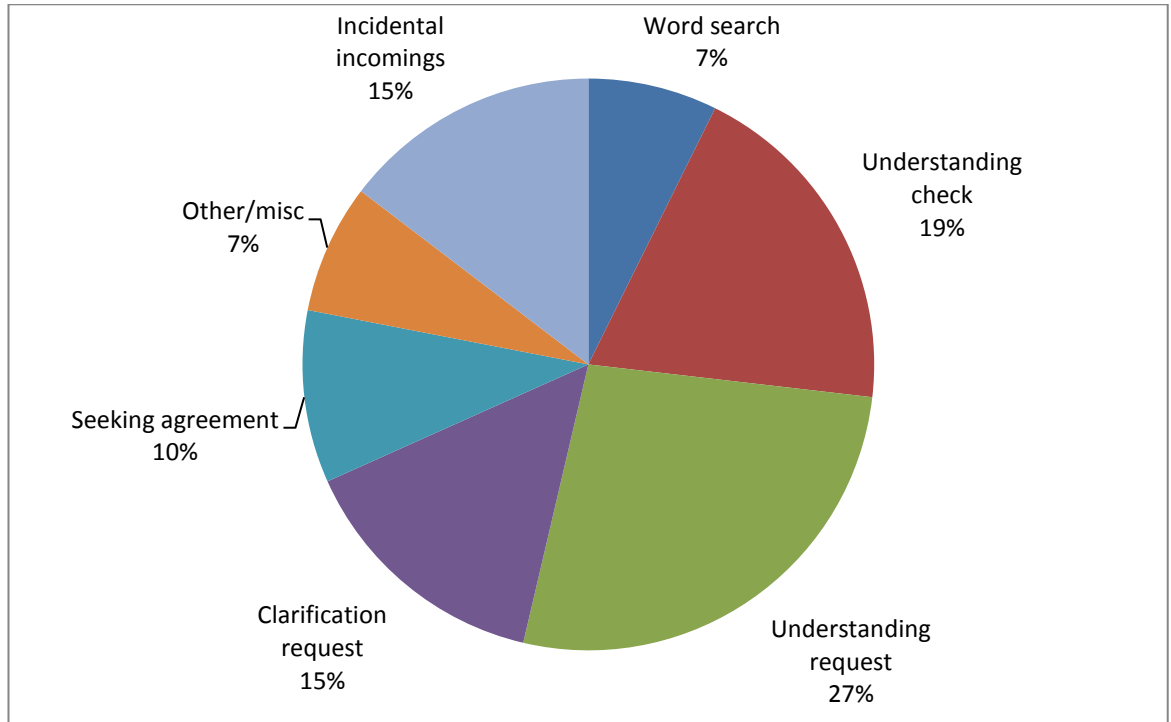

Figure 6.A. Distribution of gesture hold according to action categories.

In addition to these 41 instances, there were a further 19 instances where a speaker gestures at turn-transition, but then releases the gesture as the co-participant starts his/her next turn. Instances of gesture hold amount to 8.4% of the turn transitions in the data, whereas gesture released at turn-transition amount to 3.9%. These numbers are summarised in Table 6.B below.

Table 6.B Distribution of gesture holds into and at turn-transition.

N instances with manual gesture at turn-transition (total N= 491)	
Hold into co-participants talk	41 (8.4%)
Hold released at turn-transition	19 (3.9%)
Total gesture hold	60 (12.2%)

Table 6.B gives an indication of how common gesture holds are at turn transitions, but it does not tell us how common gesture holds are within the different action categories. To get a better sense of how gesture hold distributes within action categories all instances of each action type in the material. These were labelled according to (i) whether or not there was a turn transition, (ii) whether or not manual gesture was used, and (iii) if manual gesture was used whether or not it was held into the co-participant's turn. I will focus only on the categories *word search* (WC), *understanding check* (UC), *understanding request* (UR) and *clarification request* (CR). There were several reasons for this choice. First, these categories seemed to have a lot in common, in terms of bringing shared understanding to the surface of interaction (i.e. unlike *incidental incomings*). Also, these categories appeared relatively clear-cut compared to others such as *seeking agreement*. A quantified summary of this analysis is given in table 6.C below.

Table 6.C. Summary of gestures and gesture hold, as used within and across turns in the action categories (AC): Word search (WS), understanding check (UC), understanding request (UR) and clarification request (CR).

	Turn-transition			No turn-transition		N (AC)
	No gesture	Gesture, no hold	GESTURE HOLD (% of total AC)	No gesture	Gesture	
WS	1	1	3 (4.5%)	43	19	67
UC	7	4	8 (40%)	1	0	20
UR	15	3	11 (34.4%)	2	1	32
CR	16	1	6 (26.1%)	0	0	23
Total	39	9	28 (19.7%)	46	20	142

Several interesting findings emerge from this analysis. First, it is clear that co-participant collaboration in word searches is not very common. Only five instances of word search occur in the data, of which four are accompanied by gesture, and three accompanied by gesture hold. In the residual instances of word search (n=62) there is no turn-transition. For the three other action categories on the other hand, the distribution shows the

opposite pattern: For UC there is one incidence of no turn-transition, for UR there are three, and for CR there are none. This is supporting evidence that word searches are primarily maintained as private matters, whereas UCs, URs and CRs explicitly appeal to collaboration.

The highest proportion of gesture holds are found in UCs, amounting to 40% of all instances of this action type. Also URs and CRs have proportions of about a third (34.4%) and a quarter (26.1%), respectively. The number of gesture holds is higher than the number of gestures that are not held into the next turn, both within and across action categories (cf. table 6.B). Thus when the relevant action categories are accompanied by gesture, the gesture is more often than not held into the co-participant's turn.

6.2.4 Summary

The distributional data show that gesture holds amount to about 8% of all turn-transitions. Importantly, the occurrences of gesture hold are attributable to certain types of action, or events in talk (e.g. incidental incomings). Gesture holds are rather common in understanding checks, understanding requests and clarification requests. These actions are rather similar to each other, in that they all handle knowledge, agreement and understanding in an explicit manner, and specifically seek to resolve shared understanding at the surface of interaction. Because gesture hold is frequently found in these action types, and because they are comparable, these three action categories will be explored further in this chapter.

6.3 Gesture holds bringing shared understanding to the fore of interaction

With two examples I will demonstrate how gesture hold brings shared understanding to the surface of interaction (primarily example 6.6), and then plays a crucial part in maintaining and resolving shared understanding (primarily example 6.7). The examples show how interactants' themselves orient to the relevance of gesture holds on a moment-by-moment basis, according to the proposed sequence presented in Table 6.A (p. 149).

6.3.1 Projecting explicit understanding

In example 6.6 the target action (lines 01-03) is Tor seeking a confirmation that Lars recognises the person he has in mind. This example is representative of the category *understanding request* used above, however as we will see, closer attention to its sequential environment reveals that it can also work as a 'topic proffer' (Schegloff, 2007), designed to test whether the audience is receptive of an upcoming topic.

Prior to this example, Tor and Lars have been talking about the area around Larvik, where they both grew up, and particularly social/sports activities they both were involved in. In this context, but without any further preface, Tor produces the name of a person, Torbjørn Thorsen, in 01. Unlike the examples above, the relevant gesture gets initiated after verbal elements of a turn, in 02. This shows that we can also make sense of gestures when they are not aligned with verbal elements of a turn.

With reference to Table 6.A, the relevant lines are indicated with numbers representing the three steps in initiating (1), maintaining (2) and resolving (3) shared understanding. As in chapter 5, the analysis is presented first in a verbal-only transcript, followed by a more detailed transcript including gestural (and other) detail.

```
(6.6) KTH-NO, TL, 7:13/552 "Torbjørn Thorsen"

01     T:      1->     torbjørn ´THOR`sen.
                       name
                       Torbjørn Thorsen

02             1->     (--)

03     L:      2/3->   torbjørn `THOR´sen ja; ´HA[N kjenner jeg `go]dt.
                       name                  YES  HE   KNOW    I  WELL
                       Torbjørn Thorsen (yes), I know him well

04     T:      3->                                    [mm,            ]
                                                       mm

05                     (-)/((T:nod))

06     T:              [mm/((nod))]

07     L:              [((nod))    ]

08     T:              °pth han har jo   h[an gikk i  min <<f >`KLAS´se?>]
                           HE   HAS part HE   WENT IN MY       CLASS
                       °pth he has       he was in my class

09     L:                                [eller h°          <<f > `KJENte>]:
                                          OR                      KNEW
                                          or h° knew
```

There is no response from Lars or elaboration from Tor immediately following the person reference in 01. At this point it might not even be clear whether Tor seeks a response from Lars. That is, as Tor produces only the person reference there are at least no direct verbal indications of whether he seeks a response, or, if he seeks a response, what kind of response that would be (e.g. whether Tor simply wants recognition of this person, or whether he wants Lars to connect the name with something specific in the previous talk). A closer look at the visual elements of 01-04 shows that Tor involves a gesture to seek recognition of this person from Lars, and that Lars then picks up on. There is also mutual gaze between the participants all the way until and during Lars' response in 03, enhancing the relevance for a response.

Tor initiates his gesture towards the end of the inter-turn gap in 02, which he holds until Lars' recognition is available in 03. These gestural events are transcribed and illustrated below.

```
(6.6a) KTH-NO, TL, 7:13/552 "Torbjørn Thorsen" GESTURE ANNOTATION

01    T:            Torbjørn ´THOR`sen.
```


a   b   c

```
      STEP       1          2              3
02-04 MG(T)      //....^^x-------------------........//
02-03 L:         [(-)[(-) [Torbjørn `THOR´sen ja; ´HA[N kjenner jeg `go]dt.
                          Torbjørn Thorsen  (yes)  I know him well

02-04 T:         [(-)[(-) [                          [mm,                ]
```

Tor's gesture is a pointing gesture, and its initiation, peak and release are shown in still-shots a-c. Notice that Lars initiates his display of recognition soon after the initiation of Tor's gesture. Thus it appears that Lars picks up on Tor's gesture as a cue to display recognition. In other words, Tor's gesture helps making shared understanding an explicit issue here, i.e. something Tor needs a contribution from Lars to resolve.

Tor seeks Lars' recognition, it seems, in order to project more talk about Torbjørn Thorsen. Or at least Tor's continuation in 08 shows that he now finds it relevant to elaborate on why this person was brought up in the conversation in the first place. Also Lars displays such an orientation: In 09, in overlap with Tor, Lars modifies his claim of recognition, to *kjente*/"knew" rather than *kjenner*/"know". By doing so Lars makes clear that although he recognises the person, he does not know him well. Perhaps Lars does this to disclaim any particular knowledge about Torbjørn Thorsen, and pre-empt a (potential) failed understanding/alignment with Tor' upcoming telling. In this way, Tor's understanding request is treated as relevant not only for Lars' next turn, but for future turns as well. Thus Tor's gesture is projective in more than one way.

Having described the projective elements of Tor's gesture and its consequences for the interaction, I will now focus more locally on how Lars and Tor maintain shared understanding during Lars' response in 03. In 03 Lars repeats the referent's name, followed by the particle ja/"yes", thus his response takes a name + confirmation-particle shape. This ordering serves two important functions. First, by producing the name first, Lars orients to his own demonstrated recognition as being the primary projected element, and not simply giving Tor permission to continue, which an initial ja/"yes" might have been heard as. Thus Lars displays a preference for (demonstrated) recognition over progressivity in this case (cf. Heritage, 2007). This is to be expected as Tor projects Lars' recognition[5]. Second, compared to other instances where a recipient may repeat a word from a previous turn (e.g. when initiating repair), the following *ja*/"yes" shows that the repeat was indeed a confirmation. This does not happen in repair sequences.

Tor's gesture reflects an orientation to the relevance of both of these elements: Tor holds his gesture only for as long as it takes Lars to produce name + confirmation token (see transcript 6.6a). Tor then releases his gesture and produces a verbal response as a validation of Lars' recognition. In sum, Tor's gesture clearly contributes to bringing shared understanding to the surface of interaction, and also reflects its resolution.

6.3.2 Maintaining and resolving an explicit understanding

Example 6.7 provides further evidence for how the use of gesture hold plays an important role, not only in projecting, but in *maintaining* the relevance for co-participants' explicit contribution towards shared understanding. The target action in

---

[5] Also, this forms an interesting parallel to example 5.3 in the previous chapter, where co-participant's display of recognition was made relevant with a mid-TCU pause. No gesture holds are found in mid-TCU pauses, which further supports the claim that gesture hold makes *explicit* rather than implicit display of understanding relevant next.

example 6.7 is an understanding check, which becomes problematic as there is initially no response, followed by a disconfirmation.

Prior to this excerpt Tor and Lars are addressing attitudes towards music. In Lars' view, phenomena like 'Guitar Hero' has helped in shifting people's attitudes in favour of social qualities of performing music rather than musical ambition. In the excerpt below Lars makes an analogy to the tradition of Norwegian school bands, which, unlike Swedish school-bands, also have focussed more on the fun parts of making music and not necessarily musical quality (01-03). The reference to school-bands as being Norwegian does not seem to be entirely clear to Tor, which is what he brings to the surface with an understanding check in 08.

```
(6.7) KTH-NO, TL, 11:36 "i Sverige"

01    L:         det ´FINS   ingen: °h `STO:re krav   til at   det her s
                 IT   EXISTS NO        BIG   DEMANDS TO  THAT IT  HERE
                 there are no great expectations for it (to)

02               <<all >`NOen gang skal> sˀ n: kunne bli   ´BRA:
                      SOME TIME SHALL        COULD BECOME GOOD
                      to ever        (x-)  become good

03               eller noe  sånt;
                 OR    SOME SUCH
                 or anything

04               (-)

05    T:         °mh=
                 °mh

06    L:         =man `GJØR ´det bare.
                  ONE  DOES  IT  JUST
                  you just do it

07               (.)

08    T:    1->  <<all >ja ´HER i> `SVERige—
                       YES HERE IN SWEDEN
                 (yes) here in Sweden

09          2->  (-)

10    L:   2/3-> <<f >´NEI: i> `NOR´ge men[er `jeg;]
                       NO  IN  NORWAY MEAN   I
                       no in Norway I mean

11    T:    3->                        <<f >[´J    ]A <<all >okay.>>
                                             YES           OKAY
                                             yeah okay

12    T:         [mm,  [((THROAT))]
                 mm ((throat))
```

169

```
13    L:              [°th [med ´KORPS,]
                       WITH SCHOOL-BANDS
                       with school-bands
```

Tor's candidate understanding in 08 comes after a complete turn from Lars, in 01-03 and incremented in 06. In this turn Lars assesses *det her*/"this here" (01), referring to Norwegian school-bands. There is nothing in the production of 03 or 06 indicating that Lars projects more talk himself, rather it seems like Lars is pursuing an agreement/alignment from Tor. This is also supported by the use of gaze: Lars gazes at Tor all the way during 01-07. Tor orients to the relevance of him responding in 08, by producing an understanding check and thereby displaying that reference trouble is what has prevented him from agreeing/aligning before. Tor's candidate solution to this reference trouble is *ja her i Sverige*/"(yes) here in Sweden" (08). An interesting design feature of this understanding check is the use of the initial *ja*/"yes". One thing that the use of this item potentially shows is that Tor has some access to understanding already. I would not expect to find such a turn-initial *ja* if Tor was expressing disbelief, for example.

As he produces his candidate understanding in 08, Tor produces a gesture which he holds into the inter-turn gap in 09. Tor seeks a confirmation from Lars, and as we will see, the development of this gesture hold demonstrates both interactants' orientation to its relevance for the current process. The details of this development are illustrated in transcript 6.7a below.

(6.7a) KTH-NO, TL, 11:36 "i Sverige" GESTURE ANNOTATION 1


```
          STEP      1                2    (3)  2          3
08-10  MG(T)        .........^^^^x^^^^x-----(....)--------//
08-11  T:             ja ´HER i `SVERige— (0.3)=              [´J  ]A okay.
                    (yes) here in Sweden                      yeah okay

10     L:                                    =´NEI: i `NOR´ge men[er jeg;]
                                              no   in Norway I mean
```
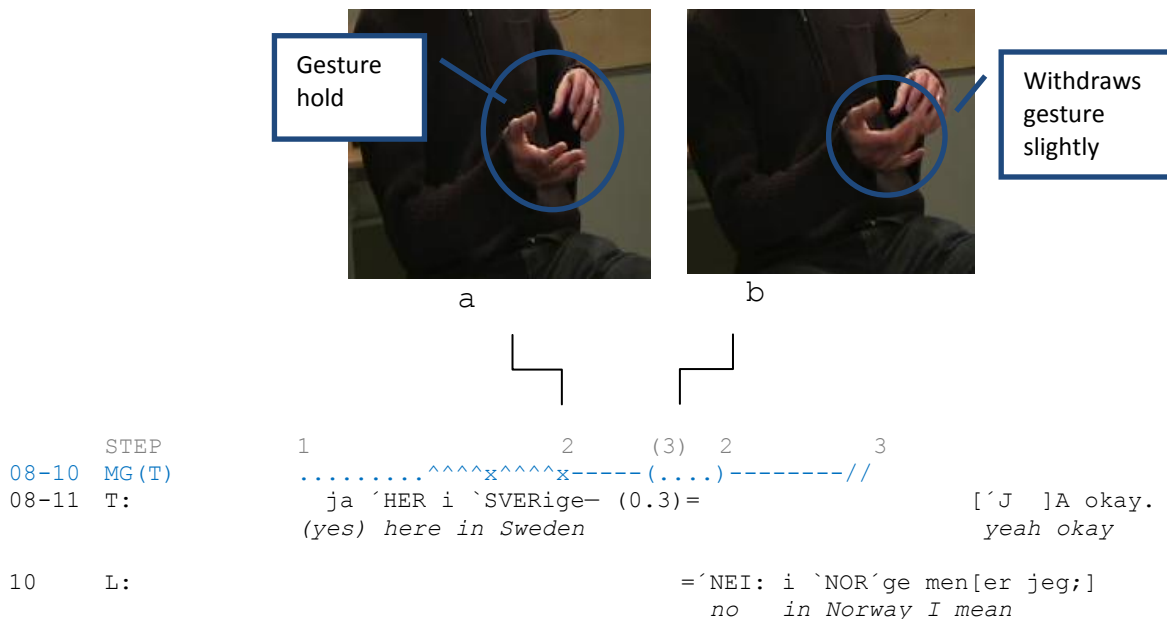
Tor holds his gesture stroke as it reaches its second peak/beat in the last syllable of
Sverige/"Sweden". This gesture is shown in figure a. For three-tenths of a second Lars
does not initiate a response. After about 2/3 of this time Tor starts to release his
gesture (figure b, and indicated by '...' in brackets). Immediately following this Lars
initiates a response, and in response, Tor holds his gesture again (i.e. the handshape in
figure b). What this shows is Tor's sensitivity to, and projection of, the emerging
contribution from Lars. As Lars does not provide any response for some time, Tor starts
withdrawing his action, but maintains it again as Lars does initiate a response. That is,
Tor holds his 'semi-released' gesture only because talk is now again aimed towards
achieving shared understanding.

But it is not only Tor who is sensitive to the gesture hold as part of the ongoing process:
As Tor starts releasing his gesture (figure b), Lars not only initiates talk, he does so in a
highly distinct and abrupt manner. First, Lars' *nei*/"no" initiating 10 is clearly louder than
the surrounding talk. Second, the *nei* is preceded and accompanied by Lars quickly
raising his shoulders. Third, Lars shifts his gaze away from Tor at the same time. The
shoulder movement ('SG') and gaze ('Gz') is included in transcript 6.7b below.

```
(6.7b) KTH-NO, TL, 11:36 "i Sverige" GESTURE, GAZE AND SHOULDER ANNOTATION

        STEP       1                 2      (*)  2            3
08-10   MG(T)      .........^^^^x^^^^x-----(....)-------//
08-11   T:         (ja) ´HEr i `Sverige— (0.3)=             [´J  ]A okay.
                   (yes) here in Sweden                      yeah okay


10      Gz(L-T)    _____B ,,L    x_____
10      SG(L)                                 ^^^
10      L:                                    =´NEI: i `NOr´ge men[er jeg;]
                                               no   in Norway  I mean
```

Tor's shoulder movement ('^^^') starts immediately after Tor starts releasing his gesture hold, along with a gaze-shift and loud speech production. Combined these signals provide a strong indication that Lars is now 'taking' speakership. As they get initiated right after Tor starts releasing his gesture, it seems highly plausible that Lars displays a sensitivity to Tor's gesture, and Lars does so for the purpose of maintaining the process towards shared understanding. In other words, Lars uses this as a 'last call' to give Tor the kind of response he projected.

So far I have mainly dealt with the fact that Lars finally does provide a contribution to Tor's understanding check; details as to how Tor and Lars resolve shared understanding have not yet been provided. As indicated by Lars' delayed response, Lars has some trouble in confirming Tor's understanding. And the reason for this is that Lars was referring to Norway, not Sweden (10). Tor's orientation to this development demonstrates how Tor not only orients to whether or not Lars responds, but also to the content of Lars' response.

As Lars produces *Norge*/"Norway" in 10, Tor starts releasing his gesture in preparation for a second gesture. The second gesture co-constructs "Norway" indexically as somewhere/something else than "Sweden". This development is illustrated in 6.7c below.

```
            STEP      2       3
10-11  MG(T)  ..)-------//.....^^^^^^^x...........//
10     L:     ´NEI: i `NOr´ge men[er jeg;]
              no    in Norway I mean

11-12  T:                        [´J     ]A okay.  mm,
                                  yeah okay mm
```
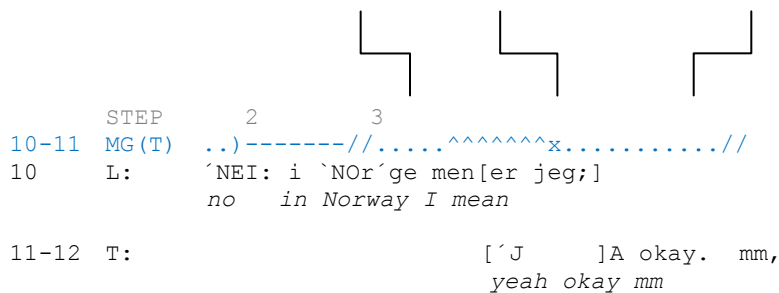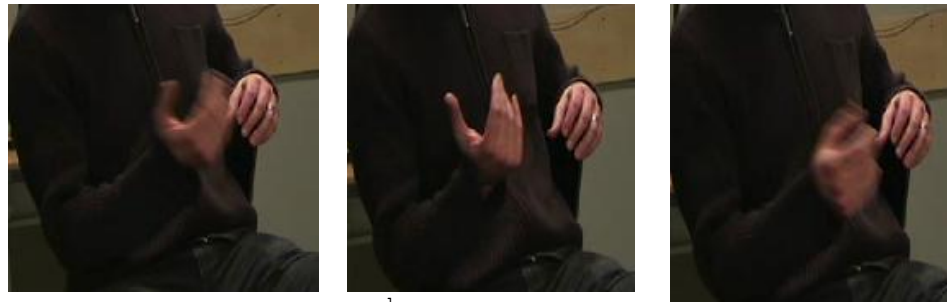
Tor's second gesture is formed as an indexical gesture using the thumb (see figures c-e above). This gesture points away from the direction of his first gesture (see transcript 6.7a), and by doing so Tor manages to anticipate, and highlight, the repair of "Sweden" to "Norway". Notice that Tor is able to display this orientation rather early, towards the coda of the first syllable in *Norge*/"Norway". There are several potential factors that may provide Tor with an opportunity to anticipate "Norway". First, as Lars has already disconfirmed Tor's understanding with *nei*/"no", the referent following *i*/"in" is likely to project another place/country. Furthermore, *i Norge*/"in Norway" uses the same lexical/syntactic format as 08, thereby marking it as the object in repair. Second, it seems plausible that the referent is either Sweden or Norway in this case, and Tor is then able to use the early parts of Lars' "Norway" to anticipate what follows.

As in 6.6, shared understanding is further resolved in a gesture release. Tor releases his gesture towards the end of Lars' TCU *i Norge mener jeg*/"in Norway I mean" (there is no good reason to view *i Norge*/"in Norway" and *mener jeg*/"I mean" as two separate TCUs, as they are produced as one intonation phrase with no phonation break), following the gesture peak (i.e. no hold). But unlike the above examples, the gesture

release is simultaneous with Tor's initiation of a verbal validation (11), and not preceding the verbal response. Note however, that Tor works to make his verbal response a next turn event (i.e. in the clear), by (i) extending *ja*/"yes" beyond the completion of Lars' spoken material in 10, and (ii) using "okay" to extend the turn with a particle that marks confirmation. Therefore, this example also supports the proposed sequence of events (Table 6.A).

6.3.3 Summary

These examples demonstrate how interactants pay attention to gestures, and gesture hold, as part of bringing and keeping shared understanding to the surface of interaction, and how they manage this process on a moment-by-moment basis. Example 6.7 in particular showed how the use of gesture is finely tuned to the resolution of shared understanding, which further proves its role and importance in projecting such an action in the first place.

6.4 Negative examples of gesture hold

Having given an overview of occurrences of gesture hold (6.2), and demonstrated the relevance of gesture hold for the achievement of shared understanding (6.3), the aim of this section is to provide an account as to when and where gesture holds work in a way that corresponds to the examples presented above. That is, when are gesture holds used, and appropriate? This question will be addressed with a set of contrastive and deviant examples, where there is either (i) no gesture hold, or (ii) gesture hold but no co-participant response. These examples will further confirm the relevance of gesture hold as seeking a co-participant's assistance in resolving an issue with understanding, but also, that mutual gaze is necessary for contextualising gesture hold as such.

The first example, 6.8 (subsection 6.4.1) will show how the use of gesture hold relates to epistemics, or who-knows-what in the interaction. More specifically, it will be argued

that presence/absence of gesture hold, along with other elements of turn design, distinguishes a claim to knowledge from suggested, or checked, knowledge. Example 6.9 (6.4.2) will give a detailed presentation of an instance where shared understanding is already accessible prior to a turn's completion. This has observable implications for the speaker's verbal/non-verbal conduct, which further supports the claims regarding the process of shared understanding. Finally, examples 6.10-6.11 (6.4.3) show how the interactional relevance of gesture holds is defined in conjunction with mutual gaze, and that the interactants negotiate co-participation accordingly.

6.4.1 Claiming knowledge

As was shown in the overview in section 6.2, there are several examples in my data where gesturing is released as co-participants initiate their turn. Such instances amount to 19 of the 60 (31.7%) of the examples I found where gesture occurs at turn-transition. In the majority of these 19 instances the gesture co-extends with the verbal content of its producer's current turn. These are then different from the examples in focus above in that a gesture is *not* maintained in orientation to the co-participant's talk, and/or the continued relevance of a projected action. This distinction relates to epistemics, and whose knowledge is relevant for the time being.

In example 6.8 below Tor claims an understanding, rather than designing the understanding as a shared project like in the examples above. In other words, Tor does not make relevant Lars' contribution to meaning in progress. As part of this process Tor releases his gesture when Lars initiates a response (lines 07-08; the relevant turn-transition is marked with '*->').

Lars has been explaining how the turntable present in the studio is used to perform musical scratching. This was initially prompted by Tor's handling the instrument. During his turn in 01-02 Lars starts handling the turntable to illustrate how the resistance is manipulated to make the scratch sounds (see figure a). *den*/"that" in 01 and 02 refers to the (moving) turntable plate (note that the intra-turn pauses in 01 are not turn breaks but periods where the interactants pay attention to visual information only). Below I

will pay primary attention to how Tor displays understanding in 03-04 and 07, in response to Lars' 01-02.

(6.8) KTH-NO, TL, 17:40 "hele tallerken"

a

```
01-02  MG(L)         ((handling turntable))
01     L:            det  vil  si  når  den `HER ´GÅR— (1.0) så: (1.5)
                     THAT WILL SAY WHEN THIS HERE GOES      THEN
                     that is to say that when this goes (1.0) then (1.5)

02                   så   kan du (.)`HOLde den;=
                     THEN CAN YOU     HOLD  IT
                     then you can (.) hold it

03     T:            =°th ja det er ´SÅNN man `GJØR <<all> ja  det er ikke man?>
                         YES IT IS  SUCH ONE  DOES        YES IT IS  NOT  ONE
                     °th yes that's how you do it        it's not you-

04                   man stopper ikke `HEle: hele  tal[´LERken] `der.
                     ONE STOPS   NOT   WHOLE WHOLE PLATEdet     THERE
                     you don't stop the entire entire record there like

05     L:                                              [nei.   ]
                                                        NO
                                                        no

06     L:            nei,=
                     NO
                     no

07     T:      *->   =liksom man ¯BAre::(m)—=
                      LIKE   ONE  JUST
                     like you just::

08     L:      *->   =man ´GJØR det n[år ma]n tar   veldig ´HARDT
                      ONE  DOES IT  WHEN ONE  TAKES VERY    HARD
                     one does it when one presses very hard

09     T:                             [mm,  ]
                                        mm

10     L:            men i[kke]: ellers.
                     BUT NOT     OTHERWISE
                     but not otherwise
```

```
11    T:                 [ja;]
                         YES
                         yes
```

Tor's first TCU *ja det er sånn man gjør det ja*/"yes that's how you do it" in 03 clearly claims rather than suggesting an understanding. Following this Tor proceeds to *demonstrate* his claimed understanding with *man stopper ikke hele tallerken*/"you don't stop the entire plate" in 04. Notice that Tor projects such a demonstration by joining the first TCU in 03 with the initiation of the next TCU (*det er ikke*/"it's not" at the end of 03).

In 07 Tor adds the increment *liksom man bare*/"like you just". This construction is accompanied by a gesture representing its predicate. That is, visualising the manner in which the user touches/handles the plate when scratching. Tor's gesture is formed by flat hand with palm facing down, which is moved as if touching the surface of the turntable plate and releasing this touch fast and lightly. This gesture is illustrated in transcript 6.8a below (still-shots b and c). Tor's gestural part of his construction continues beyond the verbal part, and completes his proposition, i.e. Tor does not provide a verbal complement to his gestural action.

(6.8a) KTH-NO, TL, 17:40 "hele tallerken" GESTURE ANNOTATION
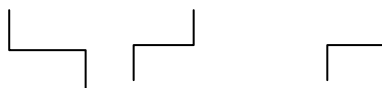

```
07-09  MG(T)      .............^^^^^^^^^^^^.....................//
07-09  T:         liksom man BAre::(m)=                [mm  ]
                  like    you just::                    mm

08     L:                                  =man ´GJør det n[år ma]n tar veldig
                                                                        ´HARdt
                                            one   does it when one presses very
                                                                        hard
```

The argument that Tor claims and demonstrates an understanding (i.e. understanding is already achieved) rather than directly appealing to Lars' co-participation for understanding to be achieved, is reflected both in Tor's use of gesture and in Lars' response. First, regarding Tor's gesture, one may observe that Tor releases his gesture (i.e. no hold) as soon as Lars initiates his response in 08 (specifically the gesture release starts during the offset of Lars' turn-initial *man*/"one"). By releasing his gesture at this point Tor shows that he no longer projects a claimed/demonstrated understanding, i.e. he does not explicitly ask for Lars' assistance to accomplish the meaning projected.

Correspondingly, in response Lars does not treat Tor's 07 as projecting a confirmation from him. Rather, Lars provides a modification of Tor's (claimed) candidate understanding as a whole. That is, in 08 (*man gjør det når man tar veldig hardt*/"one does it when one presses very hard"), *det*/"it" refers to Tor's candidate understanding of how to handle the turntable (03-04, and 07), and specifically to Tor's construction *man stopper ikke hele plata*/"you don't stop the entire plate" back in 04 (i.e. you stop the turntable plate only if you press very hard). Thus Lars' response is not specific to Tor's construction in 07, and thereby Lars shows an orientation to Tor's candidate understanding as claimed rather than 'checked'. Notice also that Lars does not initiate his response with a verbal confirmation, e.g. *ja*/"yes", thus he does not orient to the preference for dis/confirmation like in example 6.7 of understanding check above.

Thus, although shared understanding of course is an issue in example 6.8, it is not explicitly brought to the surface. Interestingly, Lars, the recipient, is the one who sits with the knowledge about the topic of talk, similar to example 6.7 above. Still, Tor's turn design is rather different in the two examples. Thus the relevance of bringing shared understanding to the surface of interaction is not necessarily about who knows most about something, but how the interactants make such distribution of knowledge relevant in their talk.

Another interesting element in this example that is yet to be addressed is the use of gaze. During his claimed understanding 03-07, Tor orients at all times towards the turn-table, and does not gaze at Lars during his turn. This lack of mutual gaze is likely to be

an additional factor in Tor maintaining his understanding as 'private' rather than shared. The use of gaze will be revisited in 6.4.3 below. First I will show how interactants orient their speech and gesture when shared meaning is achieved in the middle of a turn construction.

6.4.2 When shared understanding is already available

Above it was argued the absence of gesture hold displays a claimed rather than suggested, or checked, understanding. In a sense then, what Tor is displaying by releasing his gesture (example 6.8) is that shared understanding is already available when Lars starts talking. This brings us to the next example, which further demonstrates the interactional relevance of releasing a gesture. Here Anne abandons her own turn production (verbal and gestural) in order to display that shared understanding is already available, and she does so on the basis of Oscar's simultaneous contributions. This example is an important contribution to the data collection because it further illustrates how we are continuously sensitive to our verbal and visual actions when working towards shared understanding, including the use of gestures.

Oscar has been explaining how he finds it difficult to learn and use French, despite having attended conversational French courses in Stockholm. In 01 Anne suggests the generality of this problem: In order to learn a language properly you need to use it where it is spoken in everyday terms. 01 is a compound construction, and in overlap with 02, in 03, Oscar shows his ability to anticipate the projected completion of Anne's turn, by collaborating on the further turn production.

```
(6.9) KTH-NO, AO, 07:50 "befinne seg"

01    A:              ˘ja: h° (.) altså den `ENeste måten å  lære  seg
                      YES         THUS  THE  ONLY  WAY   TO LEARN refl.pron
                      yes the only way to learn

                      et språk   ¯SKIKkelig—
                      A LANGUAGE  PROPERLY
                      a language properly

02    A:    *->       det er jo [å     be    [`FINne seg˰      [ja,
                      THAT IS part TO   BE            refl.pron YES
                      that is to be (present)-                 yes

03    O:              <<all >[det er jo   å>[`BO i   la       [ndet ja,
                             THAT IS part TO  LIVE IN  COUNTRYdet    YES
                             that is to live in the country       yes

04    A:              [´DET er det; [det    ]=
                       THAT IS IT    IT
                       that's it     it

05    O:              [javis       [st. mm.]=
                       RIGHT
                       right          mm

06    A:              =°h det er jo  egentlig ^TULL    ↑jeg syns  at   det er
                          IT IS part ACTUALLY NONSENSE- I   THINK THAT IT   IS
                          it's really (quite) nonsense- I think it is

                      ´TULL    dette de ((...))
                       NONSENSE THAT  THEY
                       nonsense what they ((...))
```

Oscar initiates his collaboration in 03 by recycling the lexis/syntax in Anne's
construction in 02: Oscar reuses Anne's *det er jo å*/"that is to" following Anne's
compound break. In this way Oscar shows that he collaborates on Anne's projection.
Another aspect of Oscar's collaboration is illustrated by how Oscar progresses to the
main verb simultaneously with Anne (Oscar achieves this by producing this turn-
initiation slightly faster than Anne). That is, Oscar's *bo*/"live" (03), is time-aligned with
the prominent syllable in Anne's *beFINne*/"be (present)" (02). More precisely, the
release of the bilabial closure in Oscar's *bo* is simultaneous with the release of the
labiodental stricture in Anne's *beFinne* (see transcript 6.9a below). These two
simultaneous syllables are also the locations of pitch accents in Anne's and Oscar's
respective utterances. Arguably, this is an achievement, with which Oscar makes his
actions recognisable as being co-constructive with Anne's actions.

Anne:

b ə f ɪ n ə s æ̰
befinne seg-

j ɑ
ja

Oscar:

o b u ɪ l ɑ n ə j ɑ
å bo i landet ja

572.6   573.5
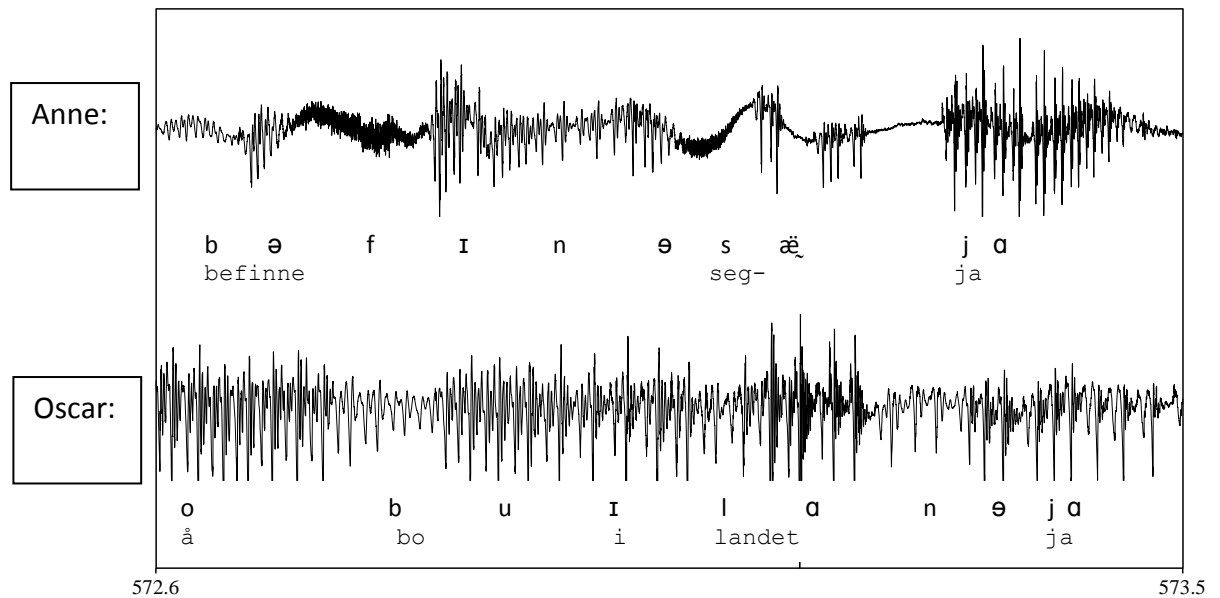
The prominent syllable of Anne's *befinne*/"be (present)" is co-expressed with a gesture, as illustrated in 6.9b.

(6.9b) KTH-NO, AO, 07:50 "befinne seg" GESTURE ANNOTATION

Withdraws
prior to
gesture
peak

```
        MG(A)                              //.......^^x^^(x)...........//
11      A:           det er jo [å      be   [`FInne  seg? [ja,
                     that is to be (present)-              yes

                     [det er jo å [`BO  i   la [ndet ja,
                     that is to live in the country yes
```

Anne's gesture appears to indexically locate a place somewhere else, by thrusting her hands in a synchronised movement away from both herself and Oscar. Figures a and b represent the main movement of this gesture. Figure b shows the peak of the gesture, which is aligned with the offset/onset between the prominent syllable and the following syllable in Anne's *befinne*/"be (present)" and Oscar's *bo*/"live". Then, as it appears that Anne is heading for another peak she withdraws her gesture (figure c and d). This happens at the same time as she halts the production of the reflexive pronoun *seg*: Anne's gestural withdrawal starts in the middle of [s] in this pronoun. Further, Anne's production of the vowel in this pronoun is strongly laryngealised (i.e. not creaky voice), as can also be seen in the waveform of x. Also, the vowel quality in *seg* is much more centralised than expected: It is realised as [sæ̈], whereas one could expect a more diphtongised [sæɪ] in most circumstances. In sum, Anne produces a combined gestural and phonatory/articulatory 'withdrawal' here.

What Anne does by withdrawing her turn production is displaying that shared understanding has been achieved, in response to Oscar having made a similar

contribution to hers. This is further confirmed by the following *ja*/"yes" (end of line 02). The fact that Anne does this and at a time when a candidate understanding is accessible from Oscar's contribution, shows that shared understanding is of fundamental relevance to her, and clearly a more central aspect of her action than completing her own proposition. It is only for this reason that Anne abandons her own verbal/gestural actions in the manner and at the time that she does.

Example 6.9 demonstrates how we orient to the implications of our own actions, in real time and while we speak. In relation to the examples of gesture holds above, example 6.9 supports the claim that gestures (and other actions that maintain an action trajectory) are only in existence for as long as shared understanding is still a relevant process. In relation to the proposed sequence (Table 6.A), Anne moves from step 1 to step 3 in the proposed sequence (Table 6.A), because step 2 is no longer relevant.

6.4.3 The role of gaze

In the descriptions above I have briefly referred to the use of gaze. It appears in all the core examples of gesture hold that the gesturing speaker gazes at the recipient, who normally gazes back. Indeed, there are no instances in my data where an incoming candidate to a word search, understanding request, clarification request, or a confirmation of an understanding check, occurs without mutual gaze being established first. This in itself indicates that mutual gaze is a key factor in framing talk for co-participation. With two examples below I will further demonstrate that during word searches, gesture holds are only projective of co-participation when accompanied by mutual gaze.

The turn production in the first example precedes the first turn in example 6.7 above, where Lars is talking about Norwegian school-bands. The co-presence of speech, gesture and gaze is illustrated in transcript 6.10 below. Notice that Lars' gaze-shift is timed with his mid-TCU pause following *men eh d:nh*/"but uh"; a pause that along with the prior hesitations clearly signals a lack of access to a word or a formulation. Notice then that Lars' gesture is held soon after (approximately 0.2 seconds). Tor does not

indicate in any way that he is going to initiate collaboration, and Lars eventually gazes back at Tor as his turn is in full progress again.

(6.10) KTH-NO, TL, 11:29 "ikke like gøy"


```
01    Gz(L)      _____,  , DR
01    MG(L)                                               ^^^^^^^^^^^^^^^^^_____
01    L:         ikke at   det alltid er like ´GØY men eh d:nh° ---(1.0)--
                 NOT  THAT IT  ALWAYS IS EQUAL FUN BUT
                 not that it's always as fun as other times but uh (1.0)

01    Gz(L)      DR              x____
01    MG(L)      --------------------
02               d: det  ´FINs   ingen ((...))
                     THERE EXISTS NO
                 th: there are no ((...))
```

This example shows that gesture hold does not necessarily project or orient to co-participation, and that mutual gaze appears to be a crucial part in contextualising a gesture hold as projecting co-participation.

Example 6.11 further strengthens this claim. It provides an interesting contrast to example 6.10, in that (i) there is mutual gaze during a word search, but (ii) there appears to be no response from co-participant. As we will see, the co-participant does indeed display willingness to respond but is unsuccessful in providing a candidate to the word search. What is most striking about this example however, is that the co-participant initiates a display of willingness to respond immediately following the speaker's gesture hold, showing that this is the crucial moment to collaborate.

Sigurd has displayed his interest (but lack of knowledge) about the presence of the turntable in the recording studio. Lars is familiar with this studio and describes how the turntable is used to calibrate the motion detector cameras present in the room (see also appendix C). Lars has just described how small reflectors are put on the turntable as part of the calibration, and from line 01 Lars describes how the cameras detect these as the turntable goes around. Particularly in lines 02-03, Lars displays trouble in finding a way to proceed with his descriptions, and the descriptions below will focus on how the interactants manage this.

```
(6.11) KTH-NO, SL, 02:26 "synkronisere"

01    L:              °h og  når  den `HER  går  rundt, (-)
                        AND WHEN THIS HERE GOES AROUND
                      °h and when this one goes around (-)

02    L:     *->      så:: tar  de  ˇINn— h°  (---)
                      THEN TAKE THEY IN
                      then:: they take in h° (---)

03    L:              °th[h  s]å  kan de   s[ynk  ]roni`SEre med; h° (-)
                           THEN CAN THEY SYNCHRONISE     WITH
                      °thh then they can   synchronise with h° (-)

04    S:                 [°tkh]             [(eh)]
                          °tkh               (uh)

05    L:              MEd [den  etter]somˀ
                      WITH THAT SINCE
                      with that one since

06    SL                 [m:ˀ       ]
                          (WITH)
                          (w:)
```

At 02, Lars provides the second part of the "when-then" compound construction initiated in 01: *når den her går rundt...*/"when this one goes around...", referring to the circular movement of the turntable. Here it appears that Lars tries to express how the cameras 'capture' information from the reflectors on the turntable. Lars displays some trouble in putting this idea into words though (notice the prolonged *så::,* and the pause following *inn*/"in"). Lars is clearly involved in a word search and it appears that by gazing at Sigurd, Lars provides a framework for co-participation. Lars gazes at Sigurd during the whole of 02-04, as shown in transcript 6.11a below.

However, Sigurd does not provide a candidate to the word search however (he is after all not the expert, Lars is). But he tries, following Lars' gesture hold, as shown in 6.11a below.

(6.11a) KTH-NO, SL, 02:26 "synkronisere" GESTURE AND GAZE ANNOTATION
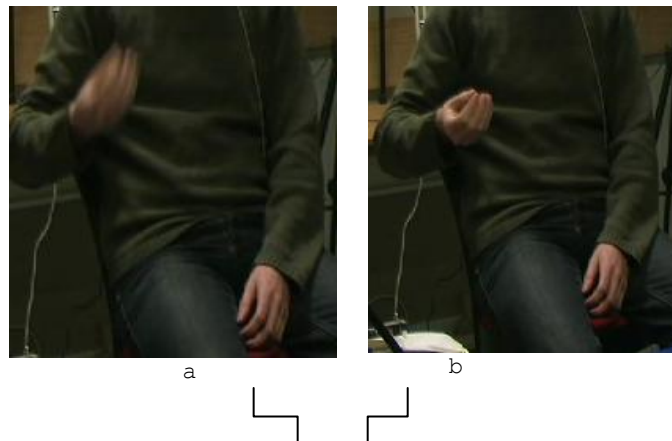

```
02-04  Gz(L)   _____
02-04  MG(L)              .....^^^^^^^--------------^^^^^^^^^^^^^^^^^^^^^^^^
02-03  L:      tar  de  ˇINn— h° (---)°th[h  s]å kan de s[ynk  ]roni`SEre
               they take in   h° (---)°thh   then they can synchronise

02-04  Gz(S)   _____,{camera},,{at turntable}
04     S:                          [°tkh]              [(eh) ]
                                   °tkh                (uh)
```

Lars introduces his gesture as a co-expression of *ta inn*/"take in". This gesture looks rather like the 'precision grip' as described by Kendon (2004), but seems iconic of 'capturing' ("take in") in this case, rather than of a more abstract 'essence' as in Kendon's data. Lars moves this hand shape vertically, during the pause following *inn* (figure a). Then, as Lars freezes this gesture in a hold (figure b), Sigurd produces an alveolar click and an inbreath 0.1 seconds later, followed by a short creaky voiced vocalic segment *(eh)* in 04. These are clear indications that Sigurd attempts to initiate a collaborative response (a further relevant observation in this regard is Sigurd's short bilabial nasal [m] in 06: It is possible that this is aimed to co-project Lars' *med...*/"with..." as it is also bilabially initiated).

The timing of Sigurd's speech sounds shows that Sigurd treats this as a relevant moment to at least display willingness to respond. Notice that this does not happen for example during Lars' 0.8 second pause in which he still produced a moving gesture: It happens exactly when Lars holds his gesture while gazing at Sigurd. In other words, Lars' gesture hold and gaze seem to trigger a response from Sigurd.

Now, as is also revealed in transcript 6.11a, Sigurd is not currently gazing at Lars as Lars holds his gesture. Sigurd has moved his gaze towards one of the cameras (I take it) during Lars' pause. However, I assume that Sigurd is equally capable of seeing Lars' gesture without looking directly at him. Furthermore, Sigurd does not just gaze away but at the instruments of which an understanding is currently made, i.e. first at one of the cameras and then at the turntable; and thereby displays that he makes a connection between the two objects in his efforts to participate.

The timing of Sigurd's speech initiation further demonstrates that there is a fine orientation to the presence of gesture hold in projection of meaning. Furthermore, examples 6.10 and 6.11 show that one semiotic resource (in this case gesture) does not work independently from other resources (in this case gaze), or from the interactional process of which it is a part.

6.4.4 Summary

The examples presented in this section enrich the understanding of how gesture holds are relevant in seeking assistance from a co-participant. First, it shows that a claim to knowledge is *not* associated with gesture hold (example 6.8). Second, it shows how the use of gesture displays sensitivity to how long a projected understanding is relevant, i.e. its use and extension depends on the moment-by-moment development of context (example 6.9). Third, it shows how gesture holds are contextualised as assistance-seeking by the accompanying cue of mutual gaze (examples 6.10-11). This study adds to the current literature on gesture, particularly by showing how gestures are used and timed according to what is relevant in the interaction. Also, it adds to the notion of co-

expressiveness the simultaneous use of gaze (see e.g. Goodwin & Goodwin, 1986; Streeck, 2009).

6.5 What extended gesture holds reveal about shared understanding

One basic argument in this study is that the interactional relevance of gesture hold is not only demonstrated in its occurrence, it is also demonstrated in its exact timing with concurrent verbal and non-verbal events (see Table 6.A, re-presented below). The relevance of this sequence was positively confirmed in examples 6.1-6.7 above. This section seeks to further demonstrate the relevance of this sequence by attending to two deviant examples. In both these examples the gesture hold is maintained during the verbal validation, thus violating step 3. Given that gesture hold is found to display a continued orientation towards shared understanding, one could expect that gesture holds that extend beyond a point where a co-participant has offered some candidate solution, display some kind of trouble with this contribution. Example 6.12 (presented in 6.5.1) will confirm this, as the extended gesture hold contextualises a concurrent verbal response as not really a confirmation. Example 6.13 (presented in 6.5.2) however, shows that extended gesture hold may also reveal a particular ownership of the candidate to shared understanding, as if the gesturing speaker had produced the candidate understanding herself.

Table 6.A – reproduced from p. 149. Formalisation of the sequence of events which lead to the achievement of shared understanding, separated in three steps and between speakers.

Speaker	Step 1	Step 2	Step 3
A	*Speaker*: Brings an issue regarding understanding to the surface of interaction (e.g. understanding check): Using verbal resources accompanied by gesture	*Speaker/hearer*: Orients to speaker B's contribution, while holding gesture	*Speaker*: Displaying achievement of shared understanding. Releasing gesture followed by verbal response
B	*Hearer*	*Hearer*: Produces contribution to shared understanding (e.g. a confirmation)	*Hearer*

6.5.1 Late gesture release displaying 'not really'

In example 6.12 Bengt tests Lars' knowledge of the Olympics, as Lars prior to the excerpt has revealed that his knowledge of this topic (and sports in general) is rather poor. In lines 02-03 Bengt projects a very specific response from Lars, namely the location of the last Olympics (Beijing in 2008). Lars produces a candidate in 04 but fails to provide the correct one, despite having been presented with the first syllable of *Peking*, i.e. Lars completes Bengt's projection as *(Pe)tersburg Sankt*. Although Bengt seems to verbally confirm Lars' following candidate (line 05), Bengt shows by extending his gesture hold that the understanding he projected is not achieved (marked with '3*' in the transcript).

```
(6.12) KTH-NO, BL, 05:10 "Peking"

01   L:              men: [eh:mˀ                 <<f >[´DA  var  det>(m)ˀ da ˀ ]
                     BUT                               THEN WAS IT       THEN
                     but   uhm                         then was it       then

02   B:                   [og  så   var det   en nå [i `SOM´mer ikke  sant  ]:=
                           AND THEN WAS THERE ON NOW IN SUMMER   NOT    TRUE
                           and then there was one this summer right

03        1->        =i: (p) (°pth)  pe::  (---)
                     IN            BEJ(JING)
                     in:           Bej::

04   L:   2->   tersburg    sa[nkt?
                (PE)TERSBURG SANKT
                tersburg    Sankt

05   B:   3*->                  [k ja: noe     [↑SÅNT?
                                 YES SOMETHING SUCH
                                 k yes something like that

06   L:                                        [`N:ET´topp, [hh° heh ]
                                                EXACTLY
                                                exactly   ((laugther))

07   B:                                                     [ja     ]
                                                             YES
                                                             yes

08   L:         jeg [har  jo  ´INGen [pˀ
                I    HAVE part NO   (CLUE)
                I don't have a c(lue)

09   B:             [(b:)            [borti `KI´na:?
                                     AWAY-IN CHINA
                                     over there in China
```

Bengt accompanies his testing action (this is an example of *understanding request*) with a gesture that is held during the pause following *Pe::*, during Lars' response in 04, and until Bengt's own verbal response *ja*/"yes" in 05. This is illustrated in 6.12a below.

(6.12a) BL, KTH-NO, 05:10 "Peking", GESTURE ANNOTATION


```
          STEP       1                    2                    3*
03-06  MG(B)    -------------------------------------//....^^^x^^^x
03-05  B:       =i: (p) (°pth) Pe:: (---)         [k ja: noe [↑SÅnt?
                in:          Bej::                 k yes something like
that

04-06  L:                                  tersburg Sa[nkt?    [`N:Ettopp,
                                           tersburg Sankt       exactly
```

Bengt's gesture is held until after the point where Lars' candidate is available, and also until after his verbal 'confirmation' *ja*/"yes". Along with the upcoming TCU, *ja noe sånt*/"yes something like that" (05), Bengt shows that he does not really validate Lars' candidate. Bengt does so by producing an unfitted response as a form of mockery towards Lars. That is, "something like that" is used as if Bengt's projection of *Peking* was not very specific, which indeed it is. This is further contextualised with Bengt's accompanying gesture, shown in figures c-d. This gesture is a shaking hand, which along with *ja noe sånt*/"something like that" appears to represent inaccuracy or uncertainty (cf. Calbris, 1990, on oscillating gestures). In sum, Bengt uses an extended gesture hold as part of displaying that the projected understanding remains unresolved.

In overlap (06), Lars designs his response in orientation to Bengt's display. In correspondence with Bengt's mockery, Lars produces a mockery 'validation' of his own candidate with a *nettopp*/"exactly" in overlap (06). That is, Lars 'validates' his own candidate with *nettopp* to signal awareness that he was *not* exactly correct. Lars displays this awareness even before Bengt has completed his mockery confirmation (in overlap with the stroke of Bengt's oscillating gesture and *sånt*), which is supporting

evidence that Lars may also attend to Bengt's continued gesture hold as an indicator of failure.

In sum, although this instance resembles some sort of 'game' between Bengt and Lars, what Bengt's extended gesture hold reveals is that shared understanding is not straightforwardly achieved, and that there are some unsuccessful elements in Lars' candidate. By extending his gesture hold during the verbal response, Bengt contextualises this response as 'not really'. This is evidenced in the way Bengt proceeds to elaborate, and how Lars aligns with Bengt's elaboration in overlap.

6.5.2 Late gesture release and ownership of candidate understanding

The final example, 6.13, is special compared to the previous examples, as it shows how interactants may not only offer a solution to each other's projected understandings, but they may work to co-construct a candidate understanding. In other words, it is not as clear as in the above examples who provides and who receives the candidate understanding/solution to an expressed issue. As in example 6.12 the gesture hold is maintained for longer than in the core examples presented above. This appears to display a claim to ownership of the successful candidate understanding, as if being the one who produced it.

Anne is in the middle of a long stretch of talk about how she had to learn several different languages as a child, and how it wasn't possible for her to retain all the different languages as her family moved from country to country. The countries her family moved between included USA, Norway, Sweden and France, and in the excerpt below she explains how she lost her ability to use French as they moved to Canada (presumably the English speaking part). Anne addresses the more general 'problem' directly in 06: *forstår du*/"do you understand", namely the consequences the constant moving had for retaining different languages.

(6.13) KTH-NO, AO 06:56 "språk"


```
01    A:              =°h og  så   da   jeg flyttet til `KANa´da? (.)
                          AND THEN WHEN I   MOVED   TO  CANADA
                       °h and then when I moved to Canada (.)

02                    og skulle:ˀ (-) (d)ta opp `ENGelsk    i´GJEN?
                      AND SHOULD      TAKE UP  ENGLISH    AGAIN
                      and were to (-) take up on English again

03                    °h °mh (pt) det gikk `FRYKTelig ´FORT?
                                  IT  WENT  TERRIBLY   QUICK
                      °h °mh (pt) it went terribly quick

04                    men da  var jeg nødt  til å `GL(h)EMm(h)e ´fransk,
                      BUT THEN WAS I  FORCED TO  TO FORGET       FRENCH
                      but then I had to              forget French

05    O:              [°pt  [ja— ]
                             YES
                              yes

06    A:     1->      [(a)  [for`ST]ÅR ´du?
                       (THUS) UNDERSTAND YOU
                       (yo-(you know)) do you understand

07                    (.)

08    O:     2->      ja.=
                      YES
                      yes

09    A:     2->      =°h altså man k `KAN ikke: °h=
                          THUS  ONE C- CAN NOT
                      °h you know one c- can't °h

10    O:     2->      =maˀ (.)[ˀ(eh) [man kan ikke] ha=
                      ONE             ONE CAN NOT   HAVE
                      on- (.)    (uh) one can't have

11    A:     2->              [det   [fˀ fiˀ ˀeh ]
                               THERE   ARE ARE
                               there a- a- uh

12    O:     2/3->    =´ALT: eh: ´LENGST `OPpi: [eh: ihˀ ] i: `HJERNen,=
                       ALL       LONGEST UP-IN    IN   IN  BRAINdet
                      everything uh furthest up in uh in- in the brain

13    A:     3->                                [nˀ nei:,]
                                                NO NO
                                                 n- no

14    O:              =[ehhˀ]
                        uh

15    A:              =[jeg ] ´TROR  at   det   finnes mennesker som `KAN ´det
                        I      THINK THAT THERE EXIST PEOPLE    THAT CAN IT
                      I think there are people who can do it

16    A:              menˀ— men jeg `KAN det ik´ke,
                      BUT   BUT I    CAN IT  NOT
                      but- but I can't
```

As Anne produces *forstår du*/"do you understand" in 06, she makes explicit that there is a main issue presented in her previous talk that has not yet been explained, and it is not yet available to mutual understanding. That is, prior to 06, this issue has only been presented implicitly, with the use of examples (moving from this place to this place, forgetting one language in favour of another, etc....). In 06 Anne projects a more definite approach to this 'problem'. This also becomes a *shared* project between Anne and Oscar as Anne produces a question-type First Pair Part: The FPP directly makes relevant a contribution from Oscar (i) by being shaped as an interrogative, (ii) by lexically addressing understanding (*forstår*/"understand"), and (iii) by addressing Oscar's understanding using the pronoun *du*/"you". Another relevant observation here is that the entire sequence from line 01 to 12 is accompanied by mutual gaze between Anne and Oscar, as a further design for sharedness.

Anne's exemplification of 'the problem' in 04 is accompanied by a manual gesture, which she repeats as she directly addresses the problem in 06. See transcript 6.13a below.

(6.13a) AO, 06:56 "språk" GESTURE ANNOTATION 1
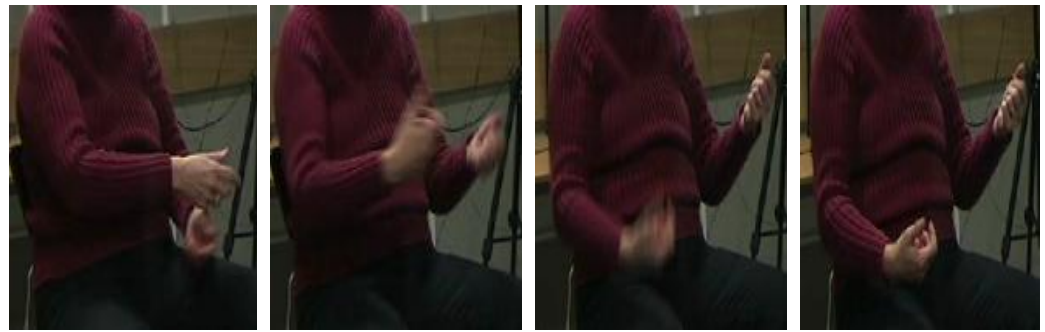
a   b   c   d

```
04    MG(A)          ...................^^^^^^^^^^^^^^^^^^x--------------
04    A:        men da var jeg nødt til å `GL(h)EMm(h)e ´fransk,
                but then  I   had to      forget       French
```

The problem: One up…

… the other one down

e   f   g
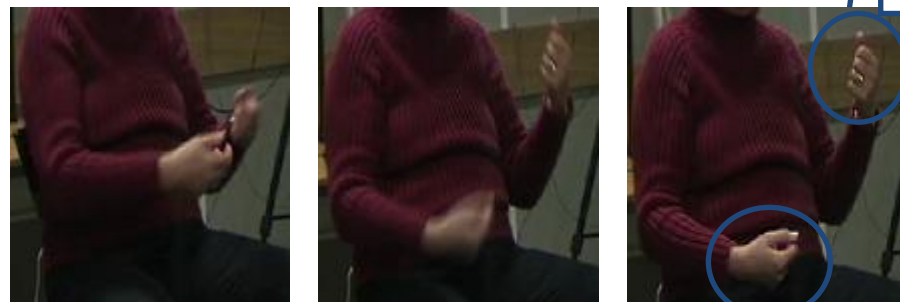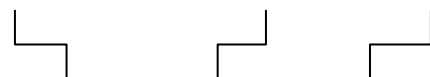
```
05-09  MG(A)     /..................^^x----------------------(---)----
05-08  O:        [°pt [ja— ]            (.) ja.
                  °pt  yes                  yes

06-09  A:   ->   [(a) [for`ST]ÅR ´du? (.)   °h altså man k `KAN ik´ke: °h
                 (*)  do you understand      °h you know one c- can't °h
```

By redoing her gesture with *forstår du/*"do you understand' (figures e-g), Anne shows
that the gesture was there as an illustration of her problem. Thus Anne accompanies
her appeal to shared understanding with an iconic representation of what the problem
she tries to get at is. Anne's hands are moving in opposite directions simultaneously,
indicating a dependent relationship between two things (i.e. two languages): Whereas
one language goes 'up', the other one goes 'down', meaning that two languages can't

be kept 'up' at the same time. Notice that the distance between the hands appears to be even greater the second time Anne produces this gesture (figure g compared to figure d), which might be a way to intensify the problem she is trying to illustrate.

Following 06 Anne holds her gesture (figure g), showing that the projected action is not yet complete. By responding with a *ja*/"yes" in 08 Oscar orients to the yes/no format of Anne's interrogative in 06. This does not sufficiently resolve shared understanding in this case, i.e. Anne's gesture hold is not about seeking confirmation, but a not yet accessible candidate understanding. Anne continues to hold her gesture while she initiates an elaboration on the understanding in 09, with *altså man kan ikke*/"you know one can't" (note that *kan ikke*/"can't" is accompanied by a small tightening of Anne's gesture, as if further locking her hands in their positions). Oscar orients to the continued relevance of his participation in 10 by co-constructing the TCU Anne initiated in 09. That is, he reuses parts of Anne's syntax/lexis in 09 and thereby co-projects the completion of Anne's turn. Anne, while holding her gesture, abandons her verbal production as Oscar proceeds in 10/12.

In other words, Anne had provided parts of a candidate solution to the problem herself, but now leaves her own attempts in favour of Oscar's. In line 12 Anne displays careful orientation to Oscar's emerging talk by mirroring his gesture. This development is illustrated in transcript 6.13b below (gesture mirroring in figures h-j).

(6.13b) AO, 06:56 "språk" GESTURE ANNOTATION 2


```
10-11  MG(A)        --------------------------------
10     O:           ma? (.)[?(eh) [man kan ikke] ha
                    on- (.)  (uh)  one  can't    have

11     A:                  [det   [f? fi? ?eh ]
                          there   a- a-   uh
```

h   i

Anne mirrors Oscar's gesture

```
12-13  MG(A)        -------------(^^^^^^^^^^^----------)..............
12-13  MG(O)         //...........^^^^x.......//
12     O:           ´ALT:   eh: ´LENGST `OPpi: [eh: ih° ] i: `HJERNen,
                    everything uh furthest up in uh  in-   in  the brain

13     A:                                      [n? nei:,]
                                               n- no
```

j
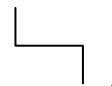
Anne holds mirrored gesture during verbal confirmation

In 12, as Oscar gets to the object *alt*/"everything", he initiates a gesture. This gesture accompanies *lengst oppi*/"everything furthest up in", which forms a significant moment in Anne's and Oscar' co-constructed candidate understanding. Oscar's gesture is shaped by lifting his hands in front of his chest, palms down and fingers facing each other. The key part of Oscar's gesture is that both his hands reach the same height-level. So accompanied by the words "one can't have everything furthest up", Oscar produces a gesture that represents two things that are at the same level. In combination, Oscar's speech and gesture could be understood as "you can't have everything up there at the same level". In response to Oscar's action/gesture, Anne lifts her hands towards the same level (represented by gesture annotation in brackets). These events lend further support to the claim made above, that Anne's visualisation of her problem is presented as 'one up, one down'. That is, as Anne lifts her hands towards the 'same level' she orients not only to Oscar's gesture, but to Oscar's solution ("you can't have everything at the same level", i.e. equally retainable) in relation to her presentation of the problem ("if one (language) is up it is at the cost of another"). Anne visibly approximates Oscar's candidate solution while he produces it, and thereby indicates that she aligns with it. Anne further confirms/validates this with the verbal response *nei*/"no" in 13.

Now, as is shown in 6.13b, Anne's gesture is still held during her verbal response *nei*/"no". This might be related to the fact that the process of bringing forward a candidate understanding is more of a shared effort in this example compared to the core examples above. In this instance, although Anne is the one who initially attempts to formulate a candidate understanding, she completes it in co-construction with Oscar. By continuing to hold her gesture, Anne may display that the candidate was exactly what she initiated but was not able to project herself. That is, she is not only confirming what Oscar has said, as recipient, but showing that she took part in producing the candidate solution as well.

6.5.3 Summary

Examples 6.12 and 6.13 show that by extending a gesture hold, the verbal response can be contextualised as non-validating (example 6.12) or as gaining rights to authorship of a gesture/candidate understanding (example 6.13). This further supports the relevance of the proposed sequence, and shows how verbal elements in talk are not understood independently of simultaneous non-verbal activity, but in close connection to them. Example 6.13 demonstrated the interactants' detailed attention to gestural meaning, as they co-construct a turn of talk. It also illustrates clearly how meaning-making is an online, and shared process.

6.6 Summary and discussion

This chapter explored the interactional resource of gesture holds as used across turn and speaker boundaries. What all instances of gesture hold into a co-participant's next turn have in common is showing that something that has been projected is not yet complete. While holding the gesture its producer displays an orientation to the co-participant's talk (thus somewhat resembling a hearer) while displaying an orientation towards the successful achievement of his/her action (thus somewhat resembling a speaker).

The study focussed on the group of activities in which gesture hold was most commonly found, namely when dealing with issues of understanding. Detailed interactional-gestural analysis showed how gesture holds are oriented to as bringing shared understanding forward and maintaining it as an explicit issue in the interaction, i.e. asking for a co-participant's assistance to resolve the issue. Second, it was demonstrated how the precise timing of the gesture hold release is oriented to as a display of resolution of shared understanding. Furthermore, a set of contrastive examples confirmed the role of gesture hold as assistance-seeking, by showing how gesture holds may be actively avoided. Also, it was shown that the gesture holds are further contextualised as assistance-seeking when there is mutual gaze between the

interactants. I would argue that the co-ordination of gesture and gaze in this way take part in forming a composite construction, and rather than talking about what qualifies as being linguistic, perhaps we ought to be more focussed on such constructions when investigating language use.

There are in particular two contributions of this research that will be addressed in this discussion:

- The relevance of studying interactional detail/timing in addressing the communicative relevance of gestures, adding for example to studies on repair
- The implications that the phenomenon of gesture hold raises for the speaker – hearer dichotomy maintained in much of research on language and interaction.

This study has not only identified a certain use of gesture that is interactional, it has also identified and demonstrated how the gestural resource makes a difference to the interactants in their management of turns at talk, by attending to how gestures 'become' interactional, i.e. how they are actively used and oriented to as the interaction emerges. Rather than simply observing *that* gestures do interactional work, this study illustrates *a particular way* in which gestures do interactional work, making a distinct contribution to what in gesture research is referred to as interactive gestures (e.g. Bavelas, 1994). Indirectly, this study shows that the 'interactiveness' of gestures is not defined by shape in particular, but by usage in relation to concurrent verbal and co-participant activities. That is, a gesture hold may be based on an iconic/representational, an indexical or a more abstract type of gesture, depending on the nature of its reference. The important aspect of it being interactional, is that it is held into a co-participant's turn.

In relation to studies on interactional uses of gesture, and CA research in particular, this study adds to others that show how gestures play a role in the design of turns at talk (e.g. Streeck & Hartge, 1992; Mondada, 2007). It shows that not only the presence of gesture, but the exact timing and co-ordination of a gesture with co-present verbal elements is important in describing its interactional use and relevance. Much of interactional research stresses the importance of investigating the unfolding moments

constituting meaning (e.g. Clark, 1996; Goodwin, 2000). This study contributes to the extant literature in demonstrating one way in which this works.

For this study it is particularly interesting to note the sequence of events which lead to a successful resolution of shared understanding. In all the core examples the gesture hold was released as soon as, and not later than the point where shared understanding was available on the basis of a co-participant's contribution (e.g. a confirmation of the gesturing speaker's candidate understanding). This release was then followed by a verbal response, and if the gesture is held during the verbal response this was not straightforwardly treated as a resolution of shared understanding. This shows that gesture not only accompanies speech as such, but contextualises verbal elements in certain ways depending on its timing with the verbal elements. Based on the well-known turn-taking constraints that only one speaker talks at a time (Sacks et al., 1974), it is not a surprise that the verbal validation follows a co-participant's verbal completion. But the finding on the timing of gesture hold may provide a better understanding of how turn-taking is managed efficiently. That is, shared understanding is already displayed as the gesture hold is released, which may provide an opportunity for a recipient to initiate a next turn in immediate overlap with a following verbal confirmation. This issue will be further addressed in the General Discussion, section 7.3.

In CA research, the most commonly addressed resource for bringing shared understanding to the surface, is repair (e.g. Schegloff et al., 1977; Drew, 1997; Curl, 2005). Although this study does not particularly address repair, many of the examples of gesture hold are indeed part of a repair initiation. Thus the findings may also be of relevance for how we understand and address the construction of repair, including non-verbal details. Also, the study highlights ways shared understanding is made an explicit issue without the use of repair. The particular actions in focus could be described as: "Do you understand what I understand?" (*understanding request*); and "is my understanding correct?" (*understanding check* and *clarification request*). What these have in common is that an explicit issue in understanding is brought to the surface, not necessarily that there is any interactional problem involved.

A further issue that this study may be used to address is what constitutes a question in talk. In previous CA research so-called 'try-marking' has been addressed as a resource for shaping a declarative as a form of query (Schegloff, 1996b), for example with the use of prosody. The examples in this study have shown that gesture holds can shape declaratives in similar ways, i.e. they contextualise an utterance as uncertain in terms of a candidate understanding, and call for co-participant assistance. It is interesting to note in this connection then that gesture may form a type of 'visual prosody', both in spoken language (e.g. Moubayed, et al., 2010) and sign language (see e.g. Dachkovsky & Sander, 2009). One might extend this study to investigate the use of gesture holds in sign language interaction.

Finally, this study shows how gestures are not restricted to the speaker's (gesturer's) own conduct, but may also co-express elements in a co-participant's talk. One interesting aspect of holding, or maintaining, a gesture is that it may extend what we traditionally think of as a complete contribution from a participant in talk. In her paper on the projection of turns with a pointing gesture, Mondada (2007) notes that: "this post-completion persistence [i.e. gesture hold] seems to contradict the idea that gestures are a speaker's characteristic and thus delimit the rights and obligations of speakership" (p. 215), implying that gestures may just as well be part of hearership as of speakership. This observation is supported in the current study. Although gesture hold in some sense displays hearership, it does not display *only* hearership, it displays a *particular kind* of combined speaker- and hearership that is inseparable from the turn preceding the hold, *and* the co-participant's contribution. In that way a gesture hold seems to display speakership and hearership at the same time. One could argue that the interactants in this way manage to display orientation to structure (i.e. 'I hold my gesture because you need to know that I am not finished, and that I am orienting to your contribution') and joint achievement (i.e. 'I hold my gesture because what you're doing is after all inseparable from what I initiated'), both at the same time.

In sum, this chapter contributes to knowledge on gesture, and particularly interactional uses of gesture. It demonstrates how gestures may initiate and maintain actions across turns of talk, and that the timing and extension of gestures with speech are crucial for how the interaction develops further. It shows that gestures' co-expressive nature is not

only about propositional meaning (cf. McNeill, 1992, 2005), but crosses speaker boundaries in online displays of interactional projects, and processes. The implications following this will be further addressed in chapter 7.

CHAPTER 7

DISCUSSION AND CONCLUSION

7.1 Main findings

This thesis was guided by three main motivations:

- Contribute with studies paying detailed attention to how hearers contribute to the interactional process
- Further potential in exploring the relation between phonetic and interactional function, as well as between non-verbal/gestural detail and interactional function
- Further potential in exploring how speakers and hearers use the interactional process to project/disambiguate what will happen next in talk, i.e. in managing turns of talk.

These motivations have been addressed in three analysis chapters, in which three phenomena were identified and studied. These involved different resources: Phonetic characteristics of verbal responses, head-nods and gesture hold. The focus was on how these resources would help contextualise current talk, and thereby define what will happen next. The findings demonstrate the detailed and precise ways in which interactants use these resources to actively take part in the management of turns and turn transitions.

The main findings were:

- Hearers affect talk-in-progress by showing how one verbal response token relates to a previous one (study 1). In particular: Hearers use phonetic characteristics to distinguish whether or not they project an uptake on current talk

- Hearers affect talk-in-progress by producing head-nods in parallel with the speaker's turn (study 2), i.e. the process of securing shared understanding depends on whether or not head-nods co-extend with the TCU to which it displays alignment

- Interactants affect talk-in-progress by holding their gesture, i.e. from their previous turn, until a co-participant response is available (study 3). The gesture holds are often part of seeking resolution to a problem, and the gesture release displays that shared understanding is resolved. While holding a gesture in this way, its producer displays features of being speaker and hearer, both at the same time.

Bringing all the analysis together, these findings show three ways in which hearers/interactants affect ongoing talk, and demonstrate that interactional meaning-making is based on speakers' and hearers' joint attention to each other and the developing talk. Perhaps more than anything else, this thesis shows that even the most routine events like completing a turn and starting another, are interactional achievements. Interactants orient to behavioural acts in relation to previous, present and upcoming events, which enable them to show that they are continuously 'up to speed' on what is related to what, and for what interactional reason. In this way the thesis contributes to previous accounts for online meaning-making and joint achievements of social action, e.g. Clark (1996), Goodwin (2000) and Mondada (2007).

The 'routine as achievement' has previously been described by Schegloff (1986), who focussed on telephone openings, and showed how even seemingly 'trivial' items like "hello" are instances of social action just like any other. This thesis shows that gestures and phonetic elements are just as 'social' as verbal elements in managing interactional routines. For example, instances of hearer responses (study 1) take particular forms and occur at particular times, in a sequence, in order to display the relation between previous, present and upcoming events. In this way hearers systematically clarify for the speaker what the connections between past and future actions are.

Much of the evidence for these findings comes from how the presence or absence of certain features affects the interactional work. This confirms that "the absence of

actions can be as decisive as their occurrence for the deployment of language and the interactional construction of discourse" Schegloff (1995, p. 186). In study 2 the presence/absence of head-nods affected what happened next in the interaction, depending on its timing with concurrent talk. However, as study 3 showed, it is not just the presence/absence of a behaviour that demonstrates the interactants' orientations to shared achievements; *altered* presence of hearer action/behaviour may do so as well, in orientation to changing circumstances. More specifically, in study 3 it was shown that hearers may change the shape of their (iconic-indexical) gesture hold according to the content of a co-participant's talk. If hearers had not altered their gestures according to the modified content, this would have been seen as absence. This shows again how interactants 'know' that which part of their presentation contributes to meaning, and how. I will argue that this temporal and sequential 'awareness' has important implications for what language is, and for how it should be studied.

As a linguistic work, this thesis shows what role interactional processes, performed in real time, play in terms of language production. The three studies show how language is shaped and accommodated according to (i) the shaping of an action, (ii) concurrent events contributing to the shaping of an action (i.e. simultaneous non-verbal/verbal productions by the same speaker, or by their co-participant), and (iii) sequence, i.e. the relation between a turn/production and a previous one. Regarding the third point, study 1 demonstrates how the relationship between the phonetic shape and function/meaning of an utterance depends on its relation to previous utterances. This finding contests the view that certain phonetic/prosodic shapes carry certain functions independently of sequential and phonetic/prosodic context, and contributes to research on the interrelationship between phonetic characteristics and interactional sequence (e.g. Curl, 2005). In coming to terms with what language is, study 1 is an example of how phonetic resources are drawn upon, in certain sequential circumstances, to disambiguate interactional options (or choices), and thus language is a set of resources *used* to perform certain actions.

Studies 2 and 3 show how interactants also shape their language productions, e.g. a word, a gesture, or an utterance, according to what a co-participant is doing in parallel, in relation to a projected meaning. These non-verbal and utterance-parallel features of

interaction has crucial relevance for whether or not speakers and hearers communicate successfully. One may want categorise them as examples of 'visual prosody' (cf. Dachkovsky & Sander, 2009; Moubayed, et al., 2010), or as non-linguisic features of communication, but in any case I would argue that they are features which make the meaningful delivery of an utterance (and linguistic units) possible (cf. Clark, 1996). The key words here are time and sequence: Language unfolds in the way it does because its producer is sensitive to how their behaviour fits with simultaneous, past and future events. I argue that linguistic science would benefit from attending to such enchronic analysis (Enfield, 2009), and from considering the relationship between a range of language resources.

In the introduction I called for an extended definition of the terms language/linguistics, as involving both verbal/spoken conduct and gestures/bodily movements, and including both conventional and non-conventional signs. I argued that a verbal or non-verbal resource qualifies as language as long as they are used systematically to project or disambiguate what action is performed, i.e. what the words and sequences of words are used *for*. Now, on the basis of the studies presented in this thesis, I would like to maintain this standpoint, which also supports previous accounts, regarding language use (Clark, 1996), mutual alignment in language modelling (Pickering & Garrod, 2004), and putting social action in the centre of linguistic analyses (Goodwin, 2000; Linell, 2009). Implications of extending our language perspective and description for future work is further addressed in section 7.3.3.

As a work on turn-taking, this thesis shows that what constitutes an interactional boundary is defined by the interactional process leading towards such a boundary. This does not mean that there are not particular cues that may commonly go along with e.g. a transition-relevance place. Rather, it means that the interactional boundary does not exist independently of the process leading towards it; and which the interactants constitute collaboratively. Studies on the organisation of turns and turn-taking should attend to such processes as well as focussing on points of transition-relevance place, and e.g. their linguistic/phonetic correlates. This thesis has sought to provide an account for how interactants understand the interactional process on a moment-by-moment basis, and how they themselves work towards an interactional boundary.

An important message that this thesis carries is that the understanding of social interaction needs attention to the details, e.g. phonetic and non-verbal, which demonstrably take part in interactants' own meaning-making.

7.2 Comments and clarification

Throughout this thesis I have used schematisations to describe the sequences of events governing the phenomenon under study (see Table 4.A, p. 75; Table 5.A, p. 112; and Table 6.A, p. 149). This was done in order to illustrate that it is a process that is explored, and to clarify what constitutes and directs that process. Also, these schematisations show that it is a generalisable sequence of events that lead to certain interactional achievements, with implications for what happens in a next turn. The illustrations below (Figure 7.A) seek to clarify how these sequences of events are based on continuous developments, over time. They show the relevant elements and the relations between them (indicated by black arrows), for each separate study. These illustrations will be referred to when evaluating to what extent the thesis has been successful in exploring the relevant aspects involved in the phenomena studied.

Study 1: Phonetic resources for 'doing the same'
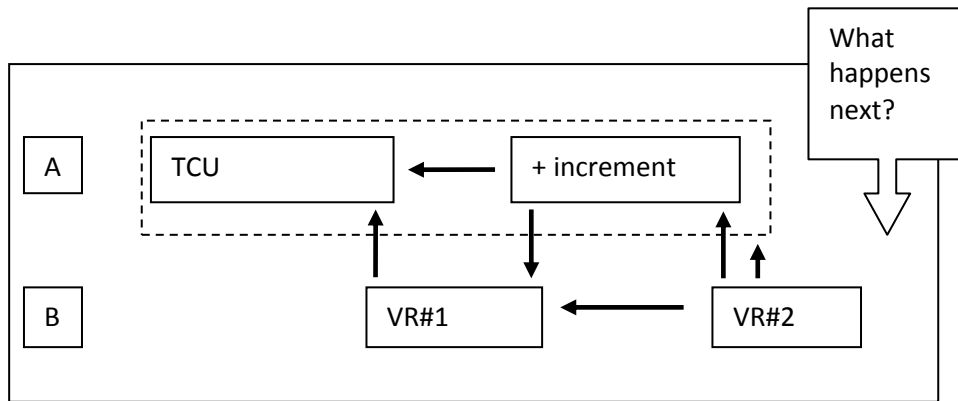
Study 2: Anticipatory nodding
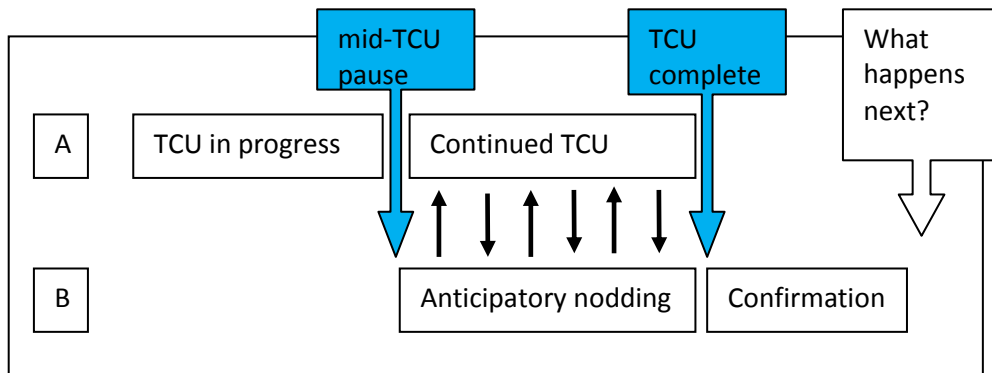
Study 3: Gesture hold
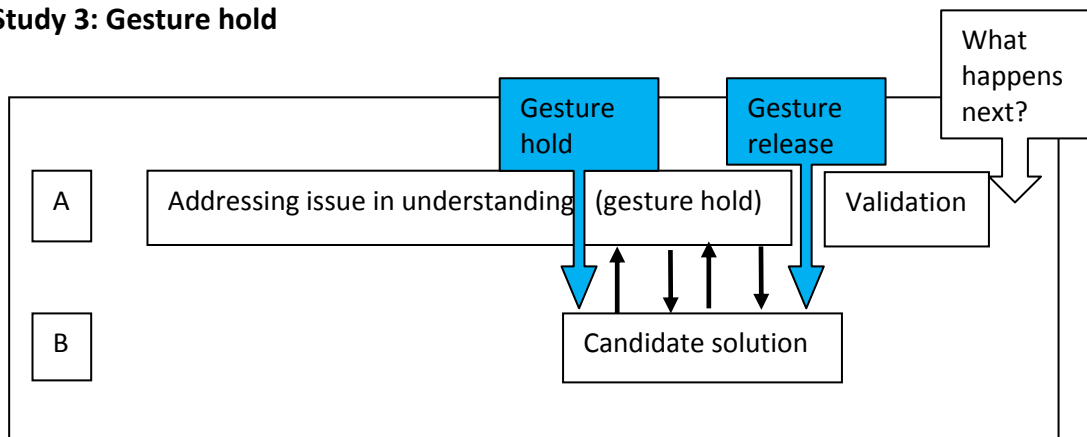
Figure 7.A. Illustrations of the sequence of events explored, in studies 1 (top), 2 (middle) and 3 (bottom). Following speaker (A/B), left to right represents time. Descriptions of the verbal/non-verbal productions presented in boxes. The black arrows represent the continuous orientation between participants (i.e. speaker A and B), highlighting the crucial relationship between their simultaneous productions (i.e. boxes). TCU = Turn-Constructional Unit, VR= Verbal Response.

With reference to continuous orientation between the interactants, different senses of the term 'continuous' emerge. In study 1 the hearer's display of continuous orientation to meaning is displayed in how a second response (VR#2) relates to a previous one (VR#1). Thus it focuses on sequential and enchronic relations, rather than continuous time *per se* (cf. Enfield, 2009). This is naturally so because the two verbal responses fill two consecutive 'slots' in the emerging interaction, i.e. they represent the two nearest points in which hearer may contribute (verbally) to the ongoing process. Studies 2 and 3 on the other hand investigate non-verbal activities which relevantly occur in parallel with a turn production, and as such, studies orientations that are truly continuous in time. Studies 2 and 3 combine the study of sequential and temporal relations, and a particular upshot of these studies is how the meaning of two verbal productions (e.g. the 'validation' and 'candidate solution' in study 3) depends on the temporal relations between the verbal conduct and the simultaneous non-verbal events. The implications this has for understanding sequential structures will be addressed further under implications in section 7.3.

Study 1 focussed mainly on the relationship between the two verbal responses. One might ask whether there are potentially further details relevant for the production of verbal responses, than were explored in this study. There is for example further potential in giving a more detailed focus on the particular uses, relevancies and shapes (e.g. prosody) of increments, in relation to the preceding TCUs (see also summary and discussion in chapter 4, section 4.5). At the same time, the phenomenon explored in this study was specifically about how a hearer is *not* displaying attention to on-going talk, when 'doing the same'. In other words, what was in focus in this study was how a hearer disregards the continuing development of talk, to show that the relation between the increment and the preceding TCU does not matter for the interactants themselves. In a sense this action 'removes' the sequential relevance of previous talk.

In sequential terms study 1 is the most complex of the studies in this thesis, because the process in focus involves two trajectories; one for the current speaker and another for the hearer, who may or may not disengage with the current talk. Studies 2 and 3 on the other hand, focussed on single actions that both interactants contribute to, in parallel.

Study 2 was based on the phenomenon of anticipatory nodding (following a mid-TCU pause) as a starting point. The importance of this study was that it showed some of the circumstances in which anticipatory nodding is relevant, when it occurs, and how its production is crucial for securing shared understanding and thereby providing a basis for continuing talking without any indication of trouble. A potential for future studies would be to explore further under what circumstances anticipatory noddings occur, i.e. when they are *not* relevant, and also what alternative means of 'doing' anticipatory alignment/understanding might look like.

Study 3 also started with a phenomenon (i.e. gesture hold), and then worked around the relevant action types in which it was found in the material. More so than study 2, this study managed to provide an overview of which actions the phenomenon was relevantly used for. This might relate to the explicit nature of this phenomenon; i.e. explicit resources might be the easiest to give an explicit account for.

Overall, the ways in which these studies were performed seem suitable for discovering new interactional phenomena, and making sense of their occurrence, and relevance. This is after all the main value of CA, the primary methodological framework for this thesis. Future research can only confirm, disconfirm and/or modify and elaborate the above findings.

For example, it would be interesting to establish to what extent these phenomena represent universal mechanisms for managing interaction, for instance, is gesture hold a publically available resource only in some cultures, or is it a surface-phenomenon displaying common principles of human interaction and cognition? I believe that the use of gesture hold displays processes that are universal to human interaction, but that the extent to which this resource, as opposed to other resources, is used may differ across cultures. It could be for example, that in some cultures (or circumstances), body movement is used rather than manual gestures to display these processes. Still, whatever the resource is, I expect to find some display of 'action not yet finished' across cultures, and that this display shows similar temporal arrangements to the one reported here.

I would argue that such potentially language-universal features are important to include in the description of a particular language. Language is, after all, a universal phenomenon, and some features involved in talk, e.g. gestures, are resources which enhance our capability to communicate cross-linguistically, and culturally (cf. Levinson, 2006). Such features might tell us something novel about the very basis for linguistic communication, and attending to them might enrich our understanding of other, more language-specific features as well.

Another issue that has been tangential to the work in this thesis has been the potential for using an experimental framework to further make advantage of the phenomena explored here. This is by no means meant to suggest that the studies in this thesis do not work well on their own, but it would be interesting to explore how, and whether, similar phenomena could be brought into lab conditions. There are obvious advantages with experimental research, which is to provide for large data-sets, and to be able to manipulate one factor at a time. Furthermore, making an effort to address interaction in an experimental framework might lead to certain methodological advances.

A further direction that future research might take is to provide quantifiable measurements of non-verbal behaviour. Besides getting more precise data, this might draw attention to further relevant detail. Technical equipment like eye-tracking and motion detectors could be used to do this. Such a study could detect details instrumentally, much like in phonetic analyses for speech. However, there are several issues involved in such a venture. First, interactional analysis depends on participants having continuous access to each other's conduct, and if the eye-tracking equipment in some ways reduces visibility, or in some way prevents interactants from gazing 'as normal', this would lead to results that are not as ecologically valid as more naturalistic data would be (due to artificiality of such interactions). Also, measuring precise movements of hands/arms might not reveal much more than is accessible to the naked eye, which, after all, is what a co-participant has access to also.

The next section will make more direct suggestions of how future research could employ experimental paradigms in continuing to explore the fundaments of talk-in-interaction. This will be done on the basis of some central implications of this thesis.

7.3 Implications and directions for future work

One conclusion to draw from this thesis is that there is a lot to be gained from investigating phonetic and gestural detail, in relation to the developing production of a turn, and sequences of turns. Indeed, the findings may serve to elaborate some of the descriptions of structures of talk-in-interaction in the literature. The implications for the particular structures of turns/TCUs, minimal sequence, and the distinction between speaker and hearer, will be addressed in subsection 7.3.1.

Further, this thesis offers insight into how interactants' display of online processes (i.e. anticipation) is used as a resource in interaction. This touches upon the interface between language, cognition and interaction (cf. Schegloff, 1991), and the implications and future directions this might lead to will be addressed in 7.3.2. Finally, section 7.3.3 provides some final notes on the potentials in this thesis for future language descriptions.

7.3.1 Non-verbal detail and structures of talk-in-interaction

TCUs, minimal sequences and turn-taking. This thesis studies how interactants behave according to structural constraints, and concepts like turn, TCU and (minimal) sequence have been used to describe these constraints. The studies show that there are a number of ways in which interactants pay attention to these structures in the way they organise their behaviour, and therefore the interactional relevance of these structures is confirmed.

For example, study 2 showed how the TCU as it is typically defined (i.e. the unit with which interaction is ordered, where syntax, prosody etc., projects its recognisable completion), is oriented to by the participants themselves as constituting a unit of action. It is only because the TCU is real for the participants that the hearer's head-nods co-extend with the speaker's turn.

Attention to non-verbal detail however, may elaborate the relevance of what happens *in between* verbal components of turns. Changes in non-verbal conduct might happen prior to any verbal response, and thereby further minimise the gap between one turn and another. In study 2 for example, an intensified nod might happen before the verbal confirmation and thus projecting this verbal confirmation. The relevance of timing of non-verbal actions in relation to verbal contributions, is clearly illustrated in study 3 (see also summary and discussion in chapter 6, section 6.6). The common structure attended to here was the minimal sequence. Instances of understanding checks for example, constituted what Schegloff (2007) refers to as minimal sequences: (i) the understanding check, (ii) the confirmation, and (iii) the validation (i.e. the verbal response of the person who checks his understanding). In Schegloff's (2007) terms these are First Pair Part (FPP), Second Pair Part (SPP) and Sequence-Closing Third (SCT), respectively. However, as was demonstrated in study 3, the meaning of the validation (SCT) depended on whether or not the gesture hold was released upon the production of the SCT. In other words, the interactional meaning of a SCT is further contextualised by what is done with the gesture. Such findings would not be available without attention to non-verbal detail, and the findings on gesture hold elaborates what is known and said about minimal sequences.

This is also relevant for understanding how efficient turn-taking is possible. That is, if a co-participant can see that a hearer aligns prior to any verbal indication of that alignment, they can prepare that for a next turn in advance. This possibility might be a specific topic of future research. For example, there are psycholinguistic techniques for exploring how early a participant makes decisions about a next turn. This way one could test how this depends on and is triggered by the particular interactional 'moments' discovered in this thesis.

Speaker – hearer dichotomy. As expected, interactants tend to produce (verbal) talk in the clear, maintaining the principle that only one speaker talks at a time (Sacks, et al., 1974). But still, there is a range of ways in which both speakers and hearers can project an upcoming turn of talk. Also, this thesis shows interactants can co-construct turns while being hearers, thus arguably showing features of being both speaker and hearer at the same time. This complicates the speaker – hearer dichotomy somewhat.

From a linear point of view (e.g. Denes & Pinson, 1993), a speaker formulates an idea using a linguistic form and expresses it with speech. Then the hearer recognises the words via the acoustic signal, parses it and reconstructs the original meaning. On the basis of this thesis, it is not at all clear how this description would fit with a hearer display of alignment and anticipation in parallel with a speaker's turn construction (study 2), or how a gesturer, who has finished speaking but maintains their gesture to show that something is not yet finished, provides meaningful information to their co-participant's talk (study 3). Thus, it does not make complete sense to investigate speakers' and hearers' roles separately: It is perhas time to consider talk as a co-ordinated process between speaker and hearer, and use categories that do not imply that one is necessarily a passive recipient (e.g. hearer) while the other is an active producer (e.g. speaker).

In terms of the last point however, it is clear that the participants are *not* symmetrical, in the sense that they contribute *in the same way* to the talk in progress. Although the speaker-hearer dichotomy is not straightforward, the interactants actively distinguish their roles and contributions, and this thesis contributes to knowledge on how these distinctions are displayed. First, in study 1, hearers show that they are hearers by producing short verbal responses, i.e. that they are not yet producing anything substantial. Also in study 2 the head-nods are clearly not used or understood to gain any speakership. In study 3, interactants produce gesture holds precisely to show that they are not currently speaking (i.e. producing a proposition), but waiting for a co-participant's contribution to confirm a proposition already made. Therefore, 'speaker' and 'hearer' (or recipient) may still be appropriate descriptions, at least for some of the different roles interactants work to distinguish, as long as one provides a detailed account on what the speaker/hearer roles are in a given circumstance.

What this boils down to though, is that speakers and hearers are constant *co-*participants, who shape and accommodate their behaviour according to each other. They are co-producers of meaning rather than simply producers and recipients.

7.3.2 Online processes as resource and constraint

In the introduction to this thesis (p. 12) I asked: 'How can one participant of a conversation make sure that the other participant is able to foresee where a turn is heading?' This thesis has answered this question by discovering some ways in which participants' display their anticipation, i.e. they make their online processes public for a co-participant to see and hear. There is major potential in exploring the systematic ways in which such anticipatory planning may work, and how interactants can access each other's online processes (cf. Goody, 1995, on 'anticipatory interactive planning'). This is the case also for language studies, or perhaps particularly so, because language is so present in our everyday life. Below I have highlighted some specific possibilities for future research.

Displaying online processes as an interactional resource. First, the thesis shows that anticipation not only reflects ability, it is also an interactional resource. For example, anticipatory nodding (study 2) is used for the interactional purpose of securing shared understanding, which may support a smooth transition from one turn to the next. Thus anticipation is not just 'there' as a pre-condition for efficient turn-taking, but is used actively to support it. Also previous research point out how anticipation is a resource. For example, Lerner's (1991, 1996) work on collaborative/anticipatory completions shows that hearers may complete a speaker's syntactic construction and in this way demonstrate their anticipation of where the rest of the speaker's talk is heading. The particular reason for doing so can be used to demonstrate agreement, or to pre-empt disagreement (Lerner, 1996).

Although they explore different phenomena, an interesting comparison emerges between Lerner's work and this thesis. Lerner (1996) attends to incomplete (or expandable) syntax as an occasion for a co-participant to contribute, and argues that a further invitation from a speaker is not required for this to happen. This might be so, but the current thesis (i.e. study 2) shows that a speaker may use certain phonetic/gestural detail to further enhance, and trigger, a hearer contribution in the middle of a turn construction. Also, this means that anticipation is not only a hearer's resource for participating in talk; a hearer's displayed anticipation may be actively

projected by speakers. In other words, there lies further potential in exploring how interactants appeal to each other's online processes, and how they may trigger co-participation in particular circumstances.

A further example is the co-ordination of gesture hold + gaze: The combination of gesture hold and mutual gaze appears to be a strong trigger effect of active co-participation toward shared understanding (see example 6.11 on pp. 175-176). That is, if a contribution to shared understanding is already somewhat relevant, what the 'frozen' gesture combined with mutual gaze does is to further elicit a particular response, i.e. disambiguating whether a co-participant collaboration is what the gesturer wants. This also fits with the more general role of mutual gaze to directly address someone (Kendon, 1977), and is a clear example of how non-verbal behaviour plays an important role in utterance construction, i.e. designing utterances to do particular things, perhaps different from utterances with similar lexical-grammatical form (cf. Enfield, 2009, on composite utterances). Knowledge of such constructions accounts for how some of the 'decision-making' in interaction is made, which is a fundamental aspect of the organisation of social action. As argued earlier in this thesis, it explains some of the 'hidden' work (cf. Heritage, 2007) that constitutes interactional management.

Interactional timing constraints. More than being a resource for managing interaction, interactants' display of online processes also forms important constraints for their conduct. An important component here is timing. Chapters 2 and 3 showed how online processes (related to shared understanding) constrain verbal and non-verbal behaviour, as displayed in their co-ordination and timing with concurrent events. Further, example 5.7 on p. 140 suggested how a speaker may increase their speech rate to accommodate a co-participant's projection of a next turn. This shows that interaction forms a powerful constraint on speech production.

The timing of interactants' conduct has also been addressed in previous CA research on turn-taking (e.g. Goodwin, 1986; Jefferson, 1986; Lerner, 1996; Schegloff, 2000). Lerner (1996) for example, reports that there are sequential limitations to when the anticipatory completions may occur, e.g. they may sound like a repetition if occurring

late, that is, after the co-participant's own completion of a turn. The fact that a production is at risk of 'sounding different' depending on its timing touches on something rather fundamental in human sense-making; namely that language production is so embedded in context that it seems impossible to tease them apart. Future research, for example within a cognitive and psycholinguistic framework, might take advantage of this and pay particular attention to timing issues that are so fundamental to language comprehension.

A specific issue for further research could be to explore *how* precise timing needs to be for a certain action to be achieved, or *whether* or not such a boundary is clear-cut. Although the current thesis conducts more precise timing measures than previous interactional research, questions about how precise these timings need to be remain open for future research. Future research could for example take advantage of the findings about gesture holds in this thesis, and explore what happens when certain non-verbal behaviour co-ordinates and does not co-ordinate with verbal content, and where the timing-boundary between different 'meanings' lies. This might be a possibility within an experimental framework, for example.

7.3.3 Implications for language description: Final notes

Above I have argued for a description of language which involves non-verbal and (non-lexical) phonetic detail, interactional structures, simultaneous speaker-hearer activities, and overall, a careful investigation of talk and interaction as it unfolds in real time.

I have argued that it is important to investigate language *constructions*, where construction refers to how speakers draw upon a range of resources in order to perform certain actions, and accommodating them to concurrent events. Such a standpoint has potential implications for future language descriptions, e.g. in terms lexicon, and grammar.

An important feature of a lexicon is that it consists of a set of decontextualised items, providing access to the function/meaning of a linguistic form without having lived an

entire life using that item. Some contextual information migh be added in a lexicon, such as whether a word form is formal/informal, and how it gets distributed according to social class. However, as we have seen in this thesis, contextualisation is more than external situation: Contextualisation happens in the way we co-ordinate different language resources. A future lexicon and grammar of language use might for example include information on how a lexical item (e.g. "okay") does certain things in certain contexts, but may in certain contexts overlap in functional terms with a "yes", provided they have certain phonetic features in common. Also, many of us are familiar with how certain grammatical structures are associated with asking a question: But there are multiple ways of marking a construction as a question, and marking the end of an utterance with a gesture hold might be one of them. The point is that it is not straightforward why such featueres of language cannot be relevant in a comprehensive language description. This thesis takes part in a growing material, within CA and interactional linguistics (e.g. Schegloff & Sacks, 1973; Lerner, 1991; Couper-Kuhlen, 1996; Schegloff, 1996a; Drew, 1997; Curl, et al., 2006), contributing to a future language description which is based on the systematic design of constructions, sensitive to interactional time and sequence.

7.3.4 Summary and conclusion

The foregoing discussion has suggested directions future research can take to further explore the 'micro-events' that are so fundamental to interactional meaning-making. In the final analysis, this thesis offers a rich account of interactional management based on micro-level observational methods. It demonstrates how the detailed study of interactional processes is central to the understanding of the way we shape and co-ordinate our behaviour and language use.

APPENDIX A

Samtyckesformulär

VR-projektet Spontal (2006-7482)

Spontal är ett svenskt projekt finansierat av Vetenskapsrådet. Det är avsett att studera mänsklig interaktion i vanliga samtal. Försökspersonerna spelas in på video samtidigt som tal och gester registreras. Allt detta sparas som datafiler.

Inspelningarna kommer endast att användas av forskare för vetenskapliga ändamål. I det ingår i första hand vetenskapliga analyser av materialet, och sammanställningar av resultaten kommer att redovisas i olika vetenskapliga rapporter, men även uppspelning av delar av materialet på exempelvis vetenskapliga konferenser i illustrativa syften kan förekomma.

Alla personuppgifter kommer att hållas helt åtskilda från inspelningarna. De kommer inte under några omständigheter att presenteras tillsammans med det inspelade materialet.

Jag har deltagit i Spontal-inspelningar och tillåter att mina inspelningar i projektet används i vetenskapliga syften enligt ovan.

Underskrift   Datum

--   --

Namnförtydligande

Consent form

VR project Spontal (2006-7482)

Spontal is a Swedish project financed by the Science Council (*Vetenskapsrådet*). It aims to study human interaction in ordinary conversation. The participants are recorded on video at the same time as speech and gestures are recorded. All of this is stored on data files.

The recordings will exclusively be used by researchers for scientific purposes. That primarily involves scientific analyses of the material, and the summary of results will be presented in different scientific journals, furthermore parts of the material may be presented at scientific conferences for illustrative purposes.

All personal information will be kept separate from the recordings. They will under no circumstances be presented along with the recorded material.

I have participated in Spontal recordings and allow the recordings of me for this project to be used for the scientific purposes as described above.

Signature   Date

Name

APPENDIX B

Explanation of translation gloss that are not direct translations.

Norwegian word	Transcription gloss	Explanation
jo	part	Literal translation is "yes". Used as a pragmatic marker for 'contrary to expectations'
så	part	This word may be used as the English "so" or "thus", and is in such cases translated accordingly. Some uses of this word however, are not so easily translatable. That is when it used as compound connector, e.g. "then", in if-then constructions
seg	refl.pron.	Reflexive pronoun (*seg*), translatable as "oneself"
Name of person, city etc. (e.g. Athen)	name	

APPENDIX C

The entries below give information about some special references in the conversations.

The turntable

The turntable present in the recording studio was used for calibrating the motion detector system (not used for these recordings). Incidentally, Lars does research on scratching, i.e. music made by physically manipulating the records on a turntable (see entry *Scratching*). The turntable is recurrently used as a point of reference for the participants, particularly for Lars and Tor, who talk about Lars' research, and for Lars and Sigurd, who talk about the motion detector system.

Scratching

Lars does research on scratching, i.e. music made by physically manipulating the records on a turntable. This becomes a topic particularly for Lars and Tor, and is enforced by the presence of the turntable (see entry above) in the recording studio.

The motion capture system

In the recording studio there are four cameras set up, in addition to those used to video these recordings). These are placed in each corner of the ceiling. These were used to record motion for participants in other recordings, but were still left intact during the recordings of KTH-NO. These become a matter of interest, especially for Lars and Sigurd. Lars is familiar with this system, and explains to Sigurd how the turntable present in the room is used to calibrate the motion detectors to detect movement. Incidentally, Lars also knows the turntable from his own research, on musical scratching (see entries *The turntable* and *Scratching*).

APPENDIX D

Tentative definitions of action categories (chapter 6), understanding request, understanding check, clarification request, and word search.

Understanding requests

- A speaker clearly appeals to a co-participant's (explicit) assistance for building *current /emerging* understanding.
- The issue is to form a specific understanding, which the action initiator cannot provide without bringing shared understanding to the surface.
- Differ from word searches in that they provide a more explicit basis for a co-participant to collaborate, and explicitly addresses him/her for co-participation

Understanding check

- Something in the previous turn that was not properly understood
- Testing own current understanding
- Not 'assertive' (as in summarising another's talk)

Clarification request

- Seeking to clarify something in the previous talk/turn to achieve understanding
- Differ from understanding checks in that a candidate understanding is not proposed

Word search

- Search for a *specific* word/formulation
- Not 'hesitation'/'thinking aloud' (e.g. in search for a next project in the interaction, or in finding a way to express 'delicate' matters)

REFERENCES

Abercrombie, D., 1964. Syllable-quantity and enclitics in English. In *Studies in phonetics and linguistics*. Oxford, UK: Oxford University Press, pp. 26-34.

Al Moubayed, S. et al., 2010. Audio-visual prosody: Perception, detection, and synthesis of prominence. In *Toward autonomous, adaptive, and context-aware multimodal interfaces: Theoretical and practical issues*. Heidelberg, Germany: Springer, pp. 55-71.

Alibali, M.W. & Goldin-Meadow, S., 1993. Modelling learning using evidence from speech and gesture. In *Proceedings of the 15th annual conference of the Cognitive Science Society*. University of Colorado, Boulder, 18-21 June 1993. Hillsdale, NJ: Lawrence Erlbaum Associates, pp. 203-208.

Auer, P., 1996. On the prosody and syntax of turn-continuations. In *Prosody in conversation*. Cambridge, UK: Cambridge University Press, pp. 57-100.

Auer, P., 2005. Projection in interaction and projection in grammar. *Text*, 25(1), pp. 7-36.

Auer, P., 2007. Why are increments such elusive objects? An afterthought. *Pragmatics*, 17(4), pp. 647-658.

Barkhuysen, P., Krahmer, E. & Swerts, M., 2006. How auditory and visual prosody is used in end-of-utterance detection. In *Proceedings for the 9th International Conference on Spoken Language Processing (Interspeech 2006 - ICSLP)*. Pittsburgh, PA, 17-21 September 2006. Pittsburgh, PA: Curran Associates, Inc., pp. 1262-1265.

Bavelas, J.B., 1994. Gestures as part of speech: Methodological implications. *Research on Language and Social Interaction*, 27(3), pp. 201-221.

Bavelas, J.B., 2007. Face-to-face dialogue as a micro-social context. In S. D. Duncan, J. Cassell, & E. T. Levy, eds. *Gesture and the dynamic dimension of language. Essays in honor of David McNeill*. Amsterdam/Philadelphia: John Benjamins, pp. 127-146.

Bavelas, J.B. et al., 1992. Interactive gestures. *Discourse Processes*, 15, pp. 469-489.

Benus, S., Gravano, A. & Hirschberg, J., 2007. The prosody of backchannels in American English. In *ICPhS XVI*. Proceedings of the 16th International congress of phonetic sciences. Saarbrücken, Germany, 6-10 August 2007. Dudweiler, Germany: Pirrot GmbH, pp. 1065-1068.

Beskow, J. et al., 2009. Project presentation: Spontal – multimodal database of spontaneous dialog. In *Proceedings of Fonetik 2008, the 11th Swedish Phonetics Conference*. University of Gothenburg, Sweden, 11-13 June 2008. Gothenburg, Sweden: Reprocentralen, pp. 190-193.

Beskow, J., Granström, B. & House, D., 2006. Visual correlates to prominence in several expressive modes. In *Proceedings of the 9th International Conference on Spoken Language Processing (Interspeech 2006 - ICSLP)*. Pittsburgh, PA, 17-21 September 2006. Pittsburgh, PA: Curran Associates, Inc., pp. 1272-1275.

Calbris, G., 1990. *Semiotics of French gesture*, Bloomington, IN: Indiana University Press.

Campbell, N., 2007. Approaches to conversational speech rhythm: Speech activity in two-person telephone dialogues. In *Proceedings for the 16th International Conference of Phonetic Sciences (ICPhS)*. Saarbrücken, Germany, 6-10 August 2007. Dudweiler, Germany: Pirrot GmbH, pp. 343-347.

Clancy, P.M. et al., 1996. The conversational use of reactive tokens in English, Japanese, and Mandarin. *Journal of pragmatics*, 26(3), pp. 355–387.

Clark, H.H., 1996. *Using language*. Cambridge, UK: Cambridge University Press.

Clark, H.H. & Brennan, S.E., 1991. Grounding in communication. In L. B. Resnick, J. M. Levine, & S. D. Teasley, eds. *Perspectives on socially shared cognition*. Washington, D.C.: American Psychological Association, pp. 127-149.

Clark, H.H. & Krych, M.A., 2004. Speaking while monitoring addressees for understanding. *Journal of Memory and Language*, 50, pp. 62-81.

Condon, W., 1976. An analysis of behavioral organization. *Sign Language Studies*, 5(13), pp. 285-318.

Condon, W.S. & Ogston, W.D., 1967. A segmentation of behavior. *Journal of Psychiatric Research*, 5, pp. 221-235.

Couper-Kuhlen, E., 1996. The prosody of repetition: On quoting and mimicry. In *Prosody in Conversation*. Cambridge, UK: Cambridge University Press, pp. 366-405.

Couper-Kuhlen, E., 2003. On initial boundary tones in English conversation. In *Proceedings for 15th International Congress of Phonetic Sciences (ICPhS)*. Universitat Autonoma de Barcelona, Spain 2003, pp. 119-122.

Couper-Kuhlen, E. & Selting, M. eds., 1996. *Prosody in conversation*. Cambridge, UK: Cambridge University Press

Cruttenden, A., 1997. *Intonation* 2nd ed., Cambridge, UK: Cambridge University Press.

Curl, T.S., 2005. Practices in other-initiated repair resolution: The phonetic differentiation of "repetitions." *Discourse Processes*, 39, pp. 1-43.

Curl, T.S., Local, J. & Walker, G., 2006. Repetition and the prosody-pragmatics interface. *Journal of Pragmatics*, 38, pp. 1721-1751.

Dachkovsky, S., & Sandler, W. (2009). Visual intonation in the prosody of a sign language. *Language and Speech*, 52, 287-314.

De Ruiter, J.P., Mitterer, H. & Enfield, N.J., 2006. Projecting the end of a speaker's turn: A cognitive cornerstone of conversation. *Language*, 82(3), pp. 515-535.

Denes, P.B. & Pinson, E.N., 1993. *The speech chain: The physics and biology of spoken language*, New York, N.Y.: W.H. Freeman.

Drew, P., 1997. "Open" class repair initiators in response to sequential sources of troubles in conversation. *Journal of Pragmatics*, 28, pp. 69-101.

Drew, P., 2005. Conversation Analysis. In K. L. Fitch & R. E. Sanders, eds. *Handbook of Language and Social Interaction*. Mahwah, NJ: Lawrence Erlbaum, pp. 71-102.

Drummond, K. & Hopper, R., 1993. Back channels revisited: Acknowledgment tokens and speakership incipiency. *Research on Language and Social Interaction*, 26(2), pp. 157-177.

Duncan, S., 1974. On the structure of speaker-auditor interaction during speaking turns. *Language in Society*, 2, pp. 161-180.

Duncan, S. & McNeill, D., 2000. Growth points in thinking-for-speaking. In D. McNeill, ed. *Language and Gesture*. Language, culture and cognition. Cambridge, UK: Cambridge University Press, pp. 141-161.

Edlund, J. & Beskow, J., 2009. Mushypeek: A framework for online investigation of audiovisual dialogue phenomena. *Language and Speech*, 52(2/3), pp. 351-367.

Edlund, J. & Heldner, M., 2005. Exploring prosody in interaction control. *Phonetica*, 62(2-4), pp. 215-226.

Edlund, J., Heldner, M. & Gustafson, J., 2005. Utterance segmentation and turn-taking in spoken dialogue systems. In B. Fisseni et al., eds. *Sprachtechnologie, mobile Kommunikation und linguistische Ressourcen*. Frankfurt am Main, Germany: Peter Lang, pp. 576-587.

Enfield, N.J., 2009. *The anatomy of meaning: Speech, gesture, and composite utterances*, Cambridge, UK: Cambridge University Press.

Ford, C.E., 2004. Contingency and units in interaction. *Discourse studies*, 6(1), pp. 27-52.

Ford, C.E., Fox, B.A. & Thompson, S.A. eds., 2002. *The language of turn and sequence*, Cambridge, UK: Cambridge University Press.

Ford, C.E. & Thompson, S.A., 1996. Interactional units in conversation: Syntactic, intonational, and pragmatic resources for the management of turns. In E. Ochs, E. A. Schegloff, & S. A. Thompson, eds. *Interaction and Grammar*. Studies in interactional sociolinguistics. Cambridge, UK: Cambridge University Press, pp. 134-184.

French, P. & Local, J., 1983. Turn-competitive incomings. *Journal of Pragmatics*, 7, pp. 17-38.

Furuyama, N., 2000. Gestural interaction between the instructor and the learner in origami instruction. In D. McNeill, ed. *Language and Gesture*. Language, culture and cognition. Cambridge, UK: Cambridge University Press.

Gardner, R., 2001. *When listeners talk: response tokens and listener stance*, Amsterdam, The Netherlands: John Benjamins.

Goldin-Meadow, S., 2003. *Hearing gesture: How our hands help us think*, Cambridge, MA: Harvard University Press.

Goodwin, C., 1979. The interactive construction of a sentence in natural conversation. In G. Psathas, ed. *Everyday Language: Studies in Ethnomethodology*. New York, NY: Irvington, pp. 97-121.

Goodwin, C., 1981. *Conversational organization: Interaction between speakers and hearers*, New York, NY: Academic Press.

Goodwin, C., 1986. Between and within: Alternative sequential treatments of continuers and assessments. *Human Studies*, 9, pp. 205-217.

Goodwin, C., 2000. Action and embodiment within situated human interaction. *Journal of Pragmatics*, 32, pp. 1489-1522.

Goodwin, C. & Goodwin, M.H., 1986. Gesture and Coparticipation in the Activity of Searching for a Word. *Semiotica*, 62(1-2), pp. 51-75.

Goodwin, C. & Goodwin, M.H., 1987. Concurrent operations on talk: Notes on the interactive organization of assessments. *IPRA Papers in Pragmatics*, 1(1), pp. 1-54.

Goodwin, C. & Goodwin, M.H., 1992. Assessments and the construction of context. In A. Duranti & C. Goodwin, eds. *Rethinking Context: Language as an Interactive Phenomenon*. Cambridge, UK: Cambridge University Press, pp. 147-190.

Goodwin, C. & Heritage, J., 1990. Conversation Analysis. *Annual review of anthropology*, 19, pp. 283-307.

Goodwin, M.H., 1980. Processes of mutual monitoring implicated in the production of description sequences. *Sociological Inquiry*, 50, pp. 303-317.

Goody, E.N. ed., 1995. *Social intelligence and interaction*, Cambridge, UK: Cambridge University Press.

Guaitella, I. et al., 2009. Are eyebrow movements linked to voice variations and turn-taking in dialogue? An experimental investigation. *Language and Speech*, 52(2/3), pp. 207-222.

Habets, B. et al., The role of synchrony and ambiguity in speech–gesture integration during comprehension. *Journal of Cognitive Neuroscience*, in press.

Hayashi, M., 2003a. *Joint utterance construction in Japanese conversation*, Amsterdam/Philadelphia: John Benjamins.

Hayashi, M., 2003b. Language and the body as resources for collaborative action: a study of word searches in Japanese conversation. *Research on Language and Social Interaction*, 36(2), pp. 109-141.

Heath, C., 1984. Talk and recipiency: sequential organization in speech and body movement. In J. Heritage & J. M. Atkinson, eds. *Structures of Social Action*. Studies in Conversation Analysis. Cambridge, UK: Cambridge University Press, pp. 247-265.

Heath, C., 1992. Gesture's discreet tasks: Multiple relevancies in visual conduct and in the contextualisation of language. In P. Auer & A. di Luzio, eds. *The contextualisation of language*. Amsterdam, The Netherlands: John Benjamins, pp. 101-128.

Heritage, J., 1984. A change of state token and aspects of its sequential placement. In *Structures of Social Action*. Cambridge, UK: Cambridge University Press, pp. 299-345.

Heritage, J., 1989. Current Developments in Conversation Analysis. In *Conversation: An interdisciplinary perspective*. Avon, UK: Multilingual Matters, pp. 21-47.

Heritage, J., 2007. Intersubjectivity and progressivity in person (and place) reference. In N. J. Enfield & T. Stivers, eds. *Person reference in interaction: Linguistic, cultural and social perspectives*. Language, culture and cognition. Cambridge, UK: Cambridge University Press, pp. 255-280.

Heritage, J. & Atkinson, J.M. eds., 1984. *Structures of social action*, Cambridge, UK: Cambridge University Press.

Heritage, J. & Sorjonen, M.-L., 1994. Constituting and maintaining activities across sequences: And-prefacing as a feature of question design. *Language in Society*, 23, pp. 1-29.

Holt, E. & Drew, Paul, 2005. Figurative Pivots: The use of figurative expressions in pivotal topic transitions. *Research on Language and Social Interaction*, 38(1), pp. 35-61.

House, D., Beskow, J. & Granström, B., 2001. Timing and interaction of visual cues for prominence in audiovisual speech perception. In *Proceedings of the 7th European Conference on Speech Communication and Technology (Eurospeech 2001)*. Aalborg, Denmark, 3-7 September, 2001, pp. 387-390.

Ishi, C.H., Ishiguro, H. & Hagita, N., 2006. Analysis of prosodic and linguistic cues of phrase finals for turn-taking and dialog acts. In *Proceedings for the 9th International Conference on Spoken Language Processing (Interspeech 2006 - ICSLP)*. Pittsburgh, PA, 17-21 September 2006. Pittsburgh, PA: Curran Associates, Inc.

Jaffe, J., 1987. Parliamentery procedure and the brain. In A.W. Siegman & S. Feldstein, eds. *Nonverbal behaviour and communication*. Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Jefferson, G., 1978. Sequential aspects of storytelling in conversation. In J. Schenkein, ed. *Studies in the organization of conversational interaction*. New York, NY: Academic Press, pp. 219-248.

Jefferson, G., 1983. On a failed hypothesis: "Conjunctionals" as overlap-vulnerable. *Tilburg papers in language and literature*, 28, pp. 1-33.

Jefferson, G., 1984. On stepwise transition from talk about a trouble to inappropriately next-positioned matters. In J. M. Atkinson & J. Heritage, eds. *Structures of social action: Studies of conversation analysis*. Cambridge, UK: Cambridge University Press, pp. 191-222.

Jefferson, G., 1985. Notes on a systematic deployment of the acknowledgement tokens "yeah" and "mmhm." *Papers in Linguistics*, 17(2), pp. 197-216.

Jefferson, G., 1986. Notes on "latency" in overlap onset. *Human Studies*, 9, pp. 153-183.

Jefferson, G., 1993. Caveat speaker: Preliminary notes on recipient topic-shift implicature. *Research on Language and Social Interaction*, 26(1), pp. 1-30.

Kendon, A., 1972. Some relationships between body motion and speech. An analysis of an example. In A. Siegman & B. Pope, eds. *Studies in dyadic communication*. Elmsford, NY: Pergamon Press, pp. 177-210.

Kendon, A., 1977. *Studies in the behavior of social interaction* T. A. Sebeok, ed., Lisse, The Netherlands: The Peter de Ridder Press.

Kendon, A., 1990. *Conducting interaction: Patterns of behaviour in focussed encounters*, Cambridge, UK: Cambridge University Press.

Kendon, A., 1994. Do gestures communicate? A review. *Research on Language and Social Interaction*, 27(3), pp. 175-200.

Kendon, A., 1995. Gestures as illocutionary and discourse structure markers in Southern Italian conversation. *Journal of Pragmatics*, 23, 247-279.

Kendon, A., 2004. *Gesture: Visible action as utterance*, Cambridge, UK: Cambridge University Press.

Kita, S., 1996. How representational gestures help speaking. In D. McNeill, ed. *Language and Gesture*. Language, culture and cognition. Cambridge, UK: Cambridge University Press, pp. 162-185.

Krauss, R.M., Morrel-Samuels, P. & Colasante, C., 1991. Do conversational gestures communicate? *Journal of Personality and Social Psychology*, 61, pp. 743-754.

Kristoffersen, G., 2000. *The phonology of Norwegian*, Oxford, UK: Oxford University Press.

Laver, J., 1994. *Principles of phonetics*, Cambridge, UK: Cambridge University Press.

Lerner, G.H., 1991. On the syntax of sentence-in-progress. *Language in Society*, 20, pp. 441-458.

Lerner, G.H., 1996. On the "semi-permeable" character of grammatical units in conversation: Conditional entry into the turn space of another speaker. In E. Ochs, E. A. Schegloff, & S. A. Thompson, eds. *Interaction and Grammar*. Cambridge, UK: Cambridge University Press, pp. 238-276.

Levinson, S., 2006. Cognition at the heart of human interaction. *Discourse Studies*, 8(1), pp. 85-93.

Liddell, S.K., 2003. *Grammar, gesture, and meaning in American Sign Language.* Cambridge, UK: Cambridge University Press.

Linell, P., 2009. *Rethinking language, mind, and world dialogically: interactional and contextual theories of human sense-making*, Charlotte, NC: Information Age Publishing.

Local, J., 2004. Getting back to prior talk: and-uh(m) as a back-connecting device in British and American English. In E. Couper-Kuhlen & C. E. Ford, eds. *Sound patterns in interaction: cross-linguistic studies of phonetics and prosody for conversation*. Amsterdam, The Netherlands: John Benjamins Publishing Company, pp. 377-400.

Local, J., 2005. On the interaction and phonetic design of collaborative completions. In W. Hardcastle & J. Beck, eds. *A Figure of Speech: a Festschrift for John Laver*. New Jersey, NY: Lawrence Erlbaum, pp. 263-282.

Local, J. & Kelly, J., 1986. Projection and "silences": Notes on phonetic and conversational structure. *Human Studies*, 9, pp. 185-204.

Local, J.K., Kelly, J. & Wells, W.H.G., 1986. Towards a phonology of conversation: turn-taking in Tyneside English. *Journal of Linguistics*, 22, pp. 411-437.

Local, J. & Walker, G., 2004. Abrupt-joins as a resource for the production of multi-unit, multi-action turns. *Journal of Pragmatics*, 36, pp. 1375-1403.

Local, J. & Walker, G., Methodological imperatives for investigating the phonetic organisation and phonological structures of spontaneous speech. *Phonetica*, 62, pp.120-130.

Loehr, D., 2007. Aspects of rhythm in gesture and speech. *Gesture*, 7(2), pp. 179-214.

Maynard, S.K., 1987. Interactional functions of a nonverbal sign. *Journal of Pragmatics*, 11, pp. 589-606.

McClave, E., 1994. Gestural beats: The rhythm hypothesis. *Journal of Psycholinguistic Research*, 23(1), pp. 45-66.

McNeill, D., 1992. *Hand and mind: What gestures reveal about thought*, Chicago, IL: The University of Chicago Press.

McNeill, D., 2005. *Gesture and thought*, Chicago, IL: The University of Chicago Press.

McNeill, D., Cassell, J. & McCullough, K.-E., 1994. Communicative effects of speech-mismatched gestures. *Research on Language and Social Interaction*, 27, pp. 223-237.

Mondada, L., 2007. Multimodal resources for turn-taking: Pointing and the emergence of possible next speakers. *Discourse Studies*, 9(2), pp. 195-226.

Moore, B.C.J., 1989. *An introduction to the psychology of hearing* 3rd ed., London, UK: Academic Press.

Müller, F.E., 1996. Affiliating and disaffiliating with continuers: Prosodic aspects of recipiency. In E. Couper-Kuhlen & M. Selting, eds. *Prosody in Conversation*. Interactional Studies. Cambridge, UK: Cambridge University Press, pp. 131-176.

Nobe, S., 2000. Where do most spontaneous representational gestures actually occur with respect to speech? In D. McNeill, ed. *Language and Gesture*. Language, Culture and Cognition. Cambridge: Cambridge University Press, pp. 186-198.

Nolan, F., 2003. Intonational equivalence: An experimental evaluation of pitch scales. *Proceedings of the 15th International Congress of Phonetic Sciences (ICPhS)*. Universitat Autonoma de Barcelona, Spain 2003, pp. 771-774.

Ochs, E., Schegloff, E.A. & Thompson, S.A. eds., 1996. *Interaction and grammar*, Cambridge, UK: Cambridge University Press.

Ogden, R., 2001. Turn transition, creak and glottal stop in Finnish talk-in-interaction. *Journal of the International Phonetics Association*, 31, pp. 139-152.

Ogden, R., 2006. Phonetics and social action in agreements and disagreements. *Journal of Pragmatics*, 38, pp. 1752-1775.

Oreström, B., 1983. *Turn-taking in English Conversation*, Lund, Sweden: LiberFörlag Ltd.

Özyurek, A., 2000. The influence of addressee location on spatial language and representational gestures of direction. In D. McNeill, ed. *Language and Gesture*. Language, culture and cognition. Cambridge, UK: Cambridge University Press, pp. 64-83.

Pickering, M.J. & Garrod, S., 2004. Toward a mechanistic psychology of dialogue. *Behavioral and Brain Sciences*, 27(02), p. 169–190.

Pomerantz, A., 1984. Agreeing and disagreeing with assessments: some features of preferred/dispreferred turn shapes. In J. M. Atkinson & J. Heritage, eds. *Structures of Social Action*. Cambridge, UK: Cambridge University Press, pp. 57-101.

Raymond, G., 2004. Prompting Action: The stand-alone "so" in ordinary conversation. *Research on Language and Social Interaction*, 37(2), pp. 185-218.

Sacks, H., 1987. On the preferences for agreement and contiguity in sequences in conversation. In G. Button & R. E. L. Lee, eds. *Talk and social organisation*. Intercommunication. Clevedon, UK: Multilingual matters Ltd, pp. 54-69.

Sacks, H., 1992. *Lectures on conversation, volumes I and II* G. Jefferson, ed., Oxford, UK: Blackwell Publishing.

Sacks, H. & Schegloff, E.A., 1979. Two preferences in the organisation of reference to persons in conversation and their interaction. In G. Psathas, ed. *Everyday Language: Studies in Ethnomethodology*. New York, NY: Irvington, pp. 15-21.

Sacks, H., Schegloff, E.A. & Jefferson, G., 1974. A simplest systematics for the organization of turn-taking for conversation. *Language*, 50, pp. 696-735.

Schegloff, E.A., 1982. Discourse as an interactional achievement: Some uses of "uh huh" and other things that come between sentences. In D. Tannen, ed. *Analyzing Discourse: Text and Talk*. Washington, DC: Georgetown University Press, pp. 71-93.

Schegloff, E.A., 1986. The routine as achievement. *Human Studies*, 9, pp. 111-151.

Schegloff, E.A., 1991. Conversation Analysis and socially shared cognition. In L. Resnick, J. Levine, & S. Teasley, eds. *Perspectives on socially shared cognition*. Washington, D.C.: American Psychological Association, pp. 150-171.

Schegloff, E.A., 1992. Repair after next turn: The last structurally provided defense of intersubjectivity in conversation. *American Journal of Sociology*, 97(5), pp. 1295-1345.

Schegloff, E.A., 1993. Reflections on quantification in the study of conversation. *Research on Language and Social Interaction*, 26(1), pp. 99-128.

Schegloff, E.A., 1995. Discourse as an interactional achievement III: The omnirelevance of action. *Research on Language and Social Interaction*, 28(2), pp. 185-211.

Schegloff, E.A., 1996a. Confirming Allusions: Toward an empirical account of action. *American Journal of Sociology*, 102(1), pp. 161-216.

Schegloff, E.A., 1996b. Turn Organization: One intersection of grammar and interaction. In E. Ochs, E. A. Schegloff, & S. A. Thompson, eds. *Interaction and Grammar*. Cambridge, UK: Cambridge University Press, pp. 52-133.

Schegloff, E.A., 1998. Reflections on studying prosody in talk-in-interaction. *Language and Speech*, 41(3-4), pp. 235-263.

Schegloff, E.A., 2000. Overlapping talk and the organization of turn-taking for conversation. *Language in Society*, 29, pp. 1-63.

Schegloff, E.A., 2007. *Sequence organization in interaction: A primer in Conversation Analysis I*, Cambridge, UK: Cambridge University Press.

Schegloff, E.A., Jefferson, G. & Sacks, H., 1977. The preference for self-correction in the organization of repair in conversation. *Language*, 53, pp. 361-382.

Schegloff, E.A. & Sacks, H., 1973. Opening up closing. *Semiotica*, 4, pp. 289-327.

Schober, M.F. & Clark, H.H., 1989. Understanding by addressees and overhearers. *Cognitive Psychology*, 21, pp. 211-232.

Selting, M., 2000. The construction of units in conversational talk. *Language in Society*, 29, pp. 477-517.

Selting, M. et al. (2009), Gesprächsanalyisches Transkriptionssystem 2 (GAT 2). *Gesprächsforschung – Online-Zeitschrift zur verbalen Interaktion*, 10. Available at: www.gespraechsforschung-ozs.de [Accessed December 5, 2010].

Sikveland, R.O., Forthcoming, 2012. Negotiating towards a next turn: Phonetics of "doing the same." *Language and Speech*.

Sikveland, R.O. et al., 2010. Spontal-N: A corpus of interactional spoken Norwegian. In *Proceedings of the 7th Conference on International Language Resources and Evaluation (LREC'10)*. Valetta, Malta, 17-23 May 2010: European Language Resources Association (ELRA), pp. 2986-2991.

Sikveland, R.O. & Ogden, R., 2010. Intersubjectivity, phonetics and gesture in the design of turns at talk. Poster presentation at Labphon 2010, Albuquerque, New Mexico, 8-10 July 2010. Available at: http://sites.google.com/site/rao1york/downloads.

Stivers, T., In press. Morality and question design: "Of course" as contesting a presupposition of askability. In T. Stivers, L. Mondada, & J. Steensig, eds. *The morality of knowing in conversation*. Cambridge, UK: Cambridge University Press.

Stivers, T., 2008. Stance, alignment, and affiliation during storytelling: When nodding is a token of affiliation. *Research on Language and Social Interaction*, 41(1), pp. 31-57.

Stivers, T. et al., 2009. Universals and cultural variation in turn-taking in conversation. In *Proceedings of the National Academy of Sciences, 106(26), pp. 10587-10592*.

Stivers, T. & Rossano, F., 2010. Mobilizing response. *Research on Language & Social Interaction*, 43(1), pp. 1-31.

Streeck, J., 1995. On projection. In E. N. Goody, ed. *Social intelligence and interaction*. Cambridge, UK: Cambridge University Press, pp. 87-110.

Streeck, J., 2009. *Gesturecraft: The manu-facture of meaning*, Amsterdam/Philadelphia: John Benjamins.

Streeck, J. & Hartge, U., 1992. Previews: Gestures at the transition place. In *Contextualization of language*. Philadelphia, PA: John Benjamins Publishing Company, pp. 135-158.

Stubbe, M., 1998. Are you listening? Cultural influences on the use of supportive verbal feedback in conversation. *Journal of Pragmatics*, 29(3), pp. 257–289.

Swerts, M. & Krahmer, E., 2008. Facial expression and prosodic prominence: Effects of modality and facial area. *Journal of Phonetics*, 36(2), pp. 219–238.

Terasaki, A., 1976. Pre-announcement sequences in conversation. *Social Science Working Paper, School of Social Science, University of California, Irvine*, 99.

Walker, G., 2004a. On some interactional and phonetic properties of increments to turns in talk-in-interaction. In E. Couper-Kuhlen & C. E. Ford, eds. *Sound Patterns in Interaction*. Amsterdam, The Netherlands: John Benjamins, pp. 147-169.

Walker, G., 2004b. *The phonetic design of turn endings, beginnings, and continuations in conversation*. PhD thesis. University of York, UK.

Ward, N., 2004. Pragmatic functions of prosodic features in non-lexical utterances. In *Proceedings of Speech Prosody 2004*. Nara, Japan, 23-26 March 2004, pp. 325–328.

Ward, N. & Tsukahara, W., 2000. Prosodic features which cue back-channel responses in English and Japanese. *Journal of Pragmatics*, 32(8), pp.1177-1207.

Yngve, V., 1970. On getting a word in edgewise. In *Papers in the 6th regional meeting of the Chicago Linguistics Society*. Chicago, IL: Chicago Linguistic Society, pp. 567-577.