The interactional accomplishment of action
Seuren, Lucas

CHAPTER 1

Introduction

> This whole book is but a draught
> – nay but the draught of a
> draught. Oh, Time, Strength,
> Cash, and Patience!
>
> ---
>
> Herman Melville, *Moby Dick*

1.1   A problem of action formation

Social action has long been recognized to be the heart of human communication; when in conversation, people are not primarily concerned with conveying meaning or information, but with doing actions (Austin, 1962; Schegloff, 1995). Even when they are conveying information such as when they are telling news or answering requests for information, people are concerned with those activities first. In order to understand the inner workings of social interaction we thus need to investigate how actions are brought off, what Schegloff (2007, p. xiv) calls the *action-formation problem*:

> How are the resources of the language, the body, the environment of the interaction, and position *in* the interaction fashioned into

> conformations designed to be, and to be recognizable by recipients
> as, particular actions – actions like requesting, inviting, granting,
> complaining, agreeing, telling, noticing, rejecting, and so on – in
> a class of unknown size?

This dissertation aims to address this problem by focusing on a very small subset of all possible actions: requests for confirmation that are implemented with declarative word order—or *Declarative Questions* in vernacular terms.[1] The problem can be characterized as follows. Well over 80% of the world's languages—781 out of 955 sampled languages—seem to have a specific sentence type for asking polar questions: a polar interrogative or yes/no interrogative. This sentence type can be designed with a question particle ($N = 585$), special verb morphology ($N = 164$), a combination of the two ($N = 15$), a specific word order ($N = 13$), or absence of a morpheme that indicates the clause is declarative ($N = 4$) (Dryer, 2013). Conventional wisdom goes that in these languages the polar interrogative is the default sentence type for asking polar questions, indeed that polar interrogatives in a sense are polar questions (see Quirk, Greenbaum, Leech, & Svartvik, 1985; Sadock, 1974; Sadock & Zwicky, 1985; Sadock, 2012).

Yet researchers have shown recurrently that in various languages that have polar interrogative syntax speakers will frequently, if not most of the time, use declarative word order to ask polar questions (e.g., Beun, 1989b; Freed, 1994; Huddleston, 1994; Stivers, Enfield, & Levinson, 2010). Consider for example the following two English utterances, the first a declarative, the second a polar interrogative (examples inspired by Collavin, 2011, p. 380):

(1)   The door is shut.

(2)   Is the door shut?

Despite their difference in word order, both (1) and (2) can be used to ask a polar question, to have the recipient (dis)confirm that the door is shut. We are thus presented with a puzzle. If languages—or more accurately, speakers of a language—have a specific sentence type for asking polar questions, why

---

[1]The label *question* is not as straightforward as it may seem: *question* is a commonsense term, not a technical one (Schegloff, 1984), and so it is not clear which actions are and which are not (declarative) questions. In this dissertation, I will use (*declarative*) *question* when discussing research in which the authors also use this term. But in my analyses I use more specific terminology, such as *request for information/confirmation* for *polar question*, and *declarative yes/no-type initiating action* or *yes/no declarative* (YND) (see G. Raymond, 2010a) for *declarative question*.

do they still use the declarative word order, the word order that is supposed to be used for assertions (see Sadock, 1974; Sadock & Zwicky, 1985; Sadock, 2012)? Or to use the terminology proposed in Schegloff's definition: if polar questions are made recognizable with the polar interrogative, how do recipients understand an utterance with declarative word order as a polar question?

1.2   Interaction as social action

Historically, the main field that has concerned itself with speech is philosophy. It has been well over sixty years since John Austin (1962) delivered his William James Lectures at Harvard University[2] in which he caused a paradigm shift in the philosophy of language by positing that speakers in social interaction are not concerned with making statements about the world that have some truth value. He argued that it is in fact generally impossible to even ascribe a truth value to most utterances that people produce. Language according to Austin is not about describing the world in a way that can be considered right or wrong, it is about *doing things*, speech acts to be specific. And those speech acts can be performed successfully or unsuccessfully—felicitously or infelicitously. These ideas led to the development of a new field in the philosophy of language in which action instead of truth value featured centrally: Speech Act Theory (hereafter, SAT) (Searle, 1969; Sbisà & Turner, 2013).

Around the same time Austin revolutionized the philosophy of language, a no less important paradigm shift took place in sociology. Harvey Sacks, who was inspired by Garfinkel's ethnomethodology (Garfinkel, 1967) and Goffman's postulate that face-to-face interaction is worthy of investigation in its own right (e.g., Goffman, 1955), began, in collaboration with Emanuel Schegloff, to investigate the moment-by-moment behavior of participants in various speech-exchange systems. One of Sacks' most far-reaching observations was that talk is ordered at very detailed levels of the interaction (see Schegloff's introduction to Sacks, 1995). It meant for the study of everyday interaction that no seemingly small detail could a priori be ruled out as having relevance for the participants in their organization and understanding of the interaction. The systematic study of talk-in-interaction that Sacks developed in collaboration with Schegloff from this observation came to be known as Conversation Analysis (hereafter, CA) (e.g., Sidnell & Stivers, 2013).

Although both SAT and CA investigate how actions in talk-in-interaction are produced and made recognizable, the methods differ fundamentally in their

---

[2]The lectures were delivered in 1955 and published posthumously in 1962.

theoretical assumptions, data, and evidence. One crucial difference between the two is that SAT as it was developed by Searle (1969) argues that actions are constituted by their *felicity conditions*. That is, SAT takes a single utterance and argues that it implements an action if the speaker has fulfilled the preconditions for that action. For example, speakers will have successfully asked a question when they lack the requested information, want to know the information, believe that the hearer possesses that information, believe that the hearer is willing to provide that information, and so forth. CA rejects this approach as inherently unsatisfactory. While a speaker by requesting information will be seen to reveal a lack of information, that revelation is an effect of implementing the request for information (see Sidnell & Enfield, 2012, 2014 on the distinction between action and effect; see also Levinson, 2013). Instead an utterance will be analyzed as implementing a question if (i) it is treated by the recipient as a question and (ii) if that uptake by the recipient is not subsequently contradicted by the speaker (Koole, 2015; Robinson, 2014; Schegloff, 1992). While felicity conditions have to be assumed to be omni-relevant, CA is interested in the verbal and embodied practices that participants use, moment by moment, to maintain an intersubjective understanding (see Schegloff's introduction in Sacks, 1995, p. xxvi; see also Enfield, 2013; Levinson, 2013; Schegloff, 1996a; Sidnell, 2013, 2014).

An additional assumption that distinguishes at least parts of SAT from CA is the Literal Force Hypothesis (LFH) (see Gazdar, 1981; Levinson, 1983):[3] the assumption that the major sentence types of a language have the illocutionairy force that is conventionally associated with them.[4] Consider again the examples given earlier:

(1) The door is shut.

(2) Is the door shut?

The approaches within SAT that embrace the LFH take the position that while both utterances have the same propositional content, they differ in their

---

[3]Gazdar (1981, p. 74) introduces a *literal meaning hypothesis*, but this term is amended by Levinson (1983). This change is likely made in light of the distinction many linguists and language philosophers make between the meaning of a sentence—its semantic content—and its illocutionary force (i.a., Frege, 1918/1956; Austin, 1962; Searle, 1969).

[4]An additional problem with the LFH is that there is no consensus on what the major sentence types are. Quirk et al. (1985) take there to be four for English: (1) declarative, (2) interrogative, (3) imperative, and (4) exclamative. Sadock and Zwicky (1985) and Levinson (1983) on the other hand treat the exclamative as a minor sentence type. If speakers rely on a form-function relationship, it is crucial to know how many such relationships there are.

illocutionairy or literal force (Collavin, 2011): (1) as it is a declarative has declarative force, while (2) as a polar interrogative has interrogative force.

That is not to say that these utterances cannot be used for other actions than making assertions or asking questions respectively, but it is not what they are designed to literally do. This means that under the LFH, when (1) is used as a question it still has declarative force, but it also has an additional, implied force: it is used to do an *indirect* speech act, a speech act where the literal force is somehow inadequate given the context.[5]

The problem with this approach according to Levinson (1983) is that most utterances would be indirect speech acts, and there does not seem to be a reason under the LFH account why this would have to be the case. An explanation has been sought in politeness, where being indirect will be understood as being polite (P. Brown & Levinson, 1987), but that would just lead us to ask why direct actions are impolite. Moreover, it is unclear how using (1) as an indirect speech act would contribute to asking a polite question. As will become clear in section 1.3 and the rest of this dissertation, participants have different concerns when designing polar questions.

CA takes a completely different perspective: When analyzing utterances we have to separate the form of the utterance from its function. That is, there is no one-to-one relation where a specific grammatical form will have a specific, invariant function that is encoded into that form (e.g., Curl, 2006; Curl & Drew, 2008; Huddleston, 1994; Levinson, 1983; Schegloff, 1984; T. Walker, 2014; G. Walker, 2017a). As Schegloff explains in the introduction to Sacks' lectures:

> The upshot of Sacks' analysis is to reject as inadequate the view that linguistic items determine the meaning or the force of an action, and to insist instead that the cultural, sequential or interactional status of the objects employed in the utterance shape the interaction of the linguistic item. (Sacks, 1995, p. xxxviii)

So all we can say is that (1) has declarative word order and (2) has polar interrogative word order.[6]

---

[5]SAT distinguishes between conventional and conversational, or Gricean, implicatures. Gordon and Lakoff (1971/1975) argue that speakers who either stated or questioned one of the felicity conditions would perform the act that is conventionally associated with that felicity condition. Searle (1975) on the other hand deals with indirect speech acts based on Grice's theory of implicature (Grice, 1975). Any indirect speech act violates on of Grice's conversational maxims, but given that the speaker will be seen to be cooperative, the implied speech act can be derived from the context. For a more extensive discussion of both theories, see Levinson (1983).

[6]The strict distinction between form and function is rarely realized in practice, as is evidenced

Because CA rejects the notion of literal meaning, it is also impossible to say what actions (1) and (2) are used for without knowing what preceded them in the interaction and what followed. Since participants in talk-in-interaction understand utterances in their (sequential) context, the action of an utterance in vacuo is simply undetermined (Wittgenstein, 1958). Utterance (2) could be understood by a recipient as doing a question, but also as a challenge or a display of disbelief, whereas (1) might be a statement, but could also be a question or a warning.[7] As analysts, we can only know what action either utterance is used to do, by studying how it is taken up (Sacks, Schegloff, & Jefferson, 1974; Schegloff, 1988b).

Note that by dropping the assumptions of literal meaning and literal force, our puzzle does not simply go away, it just takes on a different form. Instead of having to explain how declarative utterances can be understood as doing questioning, the problem becomes how any utterance gets to do questioning (Levinson, 1983; Schegloff, 1984). Given that speakers can use both declarative and polar interrogative word order to ask polar questions, the question is in which contexts do speakers use which sentence type and what do they achieve by choosing a certain type in a certain context.

This chapter

In the rest of this chapter I first discuss the methods that are used in this dissertation: Conversation Analysis and Interactional Linguistics. I provide a brief overview of the central methodological principles of CA: how participants organize turn taking and its procedural approach to intersubjectivity. These concepts serve as crucial background information not only for the analyses presented in this dissertation, but also for the discussions of the various other approaches to action formation. I subsequently summarize how CA has contributed to linguistic theory and how linguistics in turn contributes to our understanding of social interaction, focusing again on the aspect of turn taking, but also on the issue of how linguistic structures are understood to be used in the processes of action formation and action ascription. I show that instead of treating linguistic structure as invariant and similarly having an invariant meaning,

---

by the recurrent need for reminders (e.g., T. Walker, 2014; G. Walker, 2017a). Although CA does not assume that sentence types have a literal force, researchers rarely if ever take the position that both (1) and (2) require equal explanation as to how they are understood to be doing questioning (but see Schegloff, 1984).

[7] 'The way is shut. It was made by those who are Dead, and the Dead keep it, until the time comes. The way is shut.' (J.R.R. Tolkien, The Return of the King).

turns are designed to deal with local exigencies of the interaction (Mazeland, 2013), making linguistic structure not given and invariant, but emerging and even emergent (Hopper, 1987).

Following this methodological background, I discuss four approaches to the action-formation problem of what are called declarative questions or declarative requests for confirmation. All four approaches reject the LFH in a strict sense; that is, they do not presuppose that the major sentence types of a language have a literal force that determines action. But they resolve the action formation problem in different ways.

First I discuss an approach proposed by Beun (1989b). His analysis, which is grounded in Speech Act Theory, argues that in order to distinguish between declarative assertions and declarative questions participants rely on a combination of linguistic and contextual features that help to determine who is the Expert on the expressed proposition. If these features reveal that the recipient is the Expert, the declarative utterance will be understood as a question. An utterance that lacks these features can still be understood as a question if in its context of use it cannot be understood as an assertion. That is, each utterance has a preferred interpretation that can be overruled depending on where and when it is used.

The advantage of this approach is that it relies on recordings of actual conversations and its findings are thus partly grounded in participants' observable behavior. It does, however, argue for an amended version of the LFH which, as I will argue, is not feasible considering the innumerable number of actions that participants do.

Second I discuss two approaches from formal semantics: Gunlogson (2001, 2008) proposes that depending on who has what she calls *implicit authority* a declarative will be understood as a statement or a declarative question; Farkas and Roelofsen (2017) on the other hand argue that sentence types have an informative and inquisitive content (see Ciardelli, Groenendijk, & Roelofsen, 2013), and that utterances that have inquisitive content will be understood as (biased) questions.

While both can account for a broad range of cases, like Beun's proposal they cannot account for the plethora of actions we find in conversation. The proposed analyses only work for the ideal language user conceived by Chomsky (1965), where any deviation would simply have to be accounted for with some pragmatic condition. I argue therefore that these proposals could be better appreciated, if they were understood not as (universal) grammars of sentence types, but as positionally sensitive grammars (see Schegloff, 1996c).

Finally I discuss a recent proposal from CA by Heritage (2012a). In this

proposal participants distinguish between utterances that request and convey information based on their respective epistemic rights; who has primary rights to know about the addressed information. If the information falls in the domain of the speaker, an utterance will be understood as conveying information, whereas if it falls in the domain of the recipient, the utterance will be understood as requesting information/confirmation. Although this analysis has been embraced by many scholars in CA, there have been some recent criticisms (see Lynch & Macbeth, 2016a) which I briefly discuss as they pertain to the action-formation problem.

1.3   Conversation Analytic Method

This dissertation has as its aim to describe and account for how people in everyday life make use of a specific linguistic practices to understand each other and make themselves understood. It deals, in other words, with the methods by which participants make themselves *accountable* (see Garfinkel, 1967, 1968/1974). CA was developed in the 1960s to deal specifically with these issues, to develop a method of investigating actual events of daily life in a formal way (Sacks, 1984). But while language is indispensable for most forms of social interaction, it was not of itself an object of study. CA's findings, however, have had a significant impact on our understanding not just of language use, but of linguistic structure as well. So much so that over the past twenty years the investigation of linguistics in conversation has come to be a field in its own right: Interactional Linguistics (hereafter, IL). And indeed, studies in this field have shown that linguistic structure and language use cannot be as easily distinguished as some principal linguistic theories suggest.

What Sacks (1995) recurrently showed in his lectures, indeed, what he set out to show, is that in order to study, describe, and understand the norms and structures of talk-in-interaction, we do not need to first understand the mental grammar of the participants (cf. Chomsky, 1964); the "reality" of language is in fact not too complex to be described (cf. Chomsky, 1957). While it is true that conversation is rife with what one could call distractions, shifts of attentions, and errors (Chomsky, 1965, p. 3f.), these aspects of talk-in-interaction are as Sacks points out worth studying in their own right because they are in fact done in a highly organized manner.[8] In fact, while linguistic

---

[8]I do not take Chomsky's perspective here to mean that he considered linguistic performance a "trash bin" (cf. Drew, Walker, & Ogden, 2013), merely that he underestimated the degree to which performance, or talk-in-interaction, has its own order.

theories founded in Chomsky's generative approach have struggled to show underlying universalities to language (Evans & Levinson, 2009), CA and IL have shown that there are what could be called pragmatic universals, that is, interactional problems that are solved by different cultures through similar means. See for example Dingemanse et al. (2015) on universal principles of repair or Heritage (2016) on cross-linguistic regularities in the use of what are called change-of-state tokens.

In this section I first provide an introduction to CA's most central findings, and how its way of looking at talk-in-interaction allows for a unique, systematic study of language in social interaction. In doing so I motivate why this approach is suited for the questions addressed in this dissertation. I subsequently address the issues of intersubjectivity and common ground a bit more at length as they are central to the analyses in this dissertation as well as some alternative approaches that will be discussed in chapter 1.4. In closing I provide a brief overview of IL and its import for this dissertation.

1.3.1 Adjacency pairs and turn taking

CA has since its inception in the 1960s become one of the central methods for the study of social interaction. Although CA has its roots in sociology via Harold Garfinkel and Erving Goffman, and initially focused on everyday conversation (Sacks et al., 1974), it has since become an important method in various other scientific fields such as anthropology, linguistics, and psychology (see Stivers & Sidnell, 2013), and it is now also being used to study other speech-exchange systems, such as medical interaction, meetings, and interviews (see Heritage & Clayman, 2010). This broadening scope has been paramount to various real-world applications, such as preventing overprescription of antibiotics (Stivers, 2005b, 2005c, 2007), streamlining and increasing the efficacy of emergency calls (Koole & Verberg, 2017), and improving communication training (Stokoe, 2011, 2014).

In this section I discuss how CA's foundational findings in describing the "procedural infrastructure of interaction" (Schegloff, 1992, p. 1299) make possible a systematic study of talk-in-interaction. The central concepts are (i) that talk is largely organized through *adjacency pairs*—or more precisely *adjacency relationships* of which adjacency pairs are a special kind (Schegloff, 1988a, p. 113)—where some specific first action makes *conditionally relevant* a type-fitting second, and (ii) that talk is organized through a simple turn-taking system that minimizes both silences between and overlap of turns.

Already in his lectures Sacks (1995) talked about conversation as being

organized through pairs of actions. His observation was that utterances are not produced independently from one another, but that they are highly organized; a first seeking a second and seconds being produced in response to something that was hearably first. This notion was formalized as the adjacency pair:

> Adjacency pairs consist of sequences which properly have the following features: (1) two utterance length, (2) adjacent positioning of component utterances, (3) different speakers producing each utterance. (Schegloff & Sacks, 1973, p. 295)

This may seem like a rather roundabout way of stating that actions come in pairs: a first pair part (hereafter, FPP) and a second pair part (hereafter, SPP)—for example greetings and return-greetings, questions and answers, requests and grantings, and so forth. But by formalizing adjacency pairs in this manner Schegloff and Sacks (1973) opened conversation up to a manner of scientific inquiry that was simply not available before. By taking the adjacency relationship and particularly the adjacency pair as a basic unit of interaction researchers can show how participants build an interactional structure through those pairs of actions, and how coherence is achieved by an orientation to what is called "the base pairs" (Schegloff, 1990, 2007). It also makes deviations from this structure understandable not as simple statistical variations of a pattern, but as meaningful practices for the participants.

Take for example the second part of the definition: adjacent positioning of component utterances. The phrasing means that one utterance has to be provided after the other—SPPs follow FPPs—but not immediately after: things can intervene without breaking the adjacency relationship. When a recipient of an FPP subsequently produces a turn that is not recognizable as an SPP, it will generally be understood as delaying production of that SPP, and it will be "examined for its import, for what understanding should be accorded it" (Schegloff, 2007, p. 15). In other words, once a speaker has produced an FPP, anything the recipient does will be understood in relation to the adjacency pair that has been set in motion. For example, a recipient can be seen to initiate repair, signaling a problem with hearing or understanding the FPP.[9] Similarly, participants can produce sequences of talk that are subordinate to a base pair before the FPP—pre-expansion (Schegloff, 1980; Terasaki, 1976/2004)—or after the SPP—post-expansion (Davidson, 1984). And these expansions themselves are often also pair-organized (Schegloff, 1988a, 2007).

---

[9]An alternative option is a side-sequence (Jefferson, 1972) which can intervene in a larger activity or a parenthetical sequence (Mazeland, 2007) which can halt the ongoing production of a turn.

These adjacency pairs do not arise accidentally of course, and neither is providing the SPP optional. By implementing a specific type of FPP a speaker makes *conditionally relevant* an SPP (Schegloff, 1968). Upon completion of some first action the addressed recipient should normatively provide a type-fitting response. If that response is not forthcoming, that is, if the recipient takes too long in providing uptake, the absence of a response is noticeable and will be understood as the relevant non-production of the projected uptake. Although there is no fixed time limit for when a silence is understood as relevant non-production, the cut-off point has been found to lie around 700ms (Kendrick & Torreira, 2015), but it is contingent on the situation and the speed of the conversation. If conversationalists are involved in some other activity than just conversation, silences longer than 700ms may be unproblematic, but if turns are produced in quick succession a silence of 300ms may be understood as too long.

In addition to the adjacency relationship and conditional relevance, we need another pillar through which participants build up the structure of interaction: after the completion of each turn participants have to solve the problem of "who speaks next." It should be obvious that participants generally talk one after another, that silences between turns and overlap of turns are infrequent and short-lived (Stivers et al., 2009 showed that this holds in a variety of cultures), and that participants accomplish all this without having to agree in advance who can say what at which point in the conversation (Sacks et al., 1974, p. 700).

Sacks et al. (1974) showed that participants solve all these problems with a very simple turn-taking system that not only accounts for how turns are allocated moment-by-moment, but also how they are constructed. Any turn is built using a limited set of linguistic resources that are language specific. These unit-types need to meet the criterion of projectability, meaning that through these unit-types recipients can project the point at which the turn will come to possible completion. Additionally any turn can, but need not, contain a turn-allocation component, a component with which the speaker selects a specific recipient to speak next. Such a component can be obvious, like the action instantiated—when speakers in dyadic conversation produces an FPP, they thereby select the recipient to provide an SPP—or an address term, but it can also be more subtle such as gaze (e.g., Auer, 2017; Lerner, 2003; Rossano, 2013). These two components—turn-construction and turn-allocation—combined with the following set of rules give the turn-taking system for conversation:

(1) For any turn, at the initial transition-relevance place of an

initial turn-constructional unit:

(a) If the turn-so-far is so constructed as to involve the use a 'current speaker selects next' technique, then the party so selected has the right and is obliged to take next turn to speak; no others have such rights or obligations, and transfer occurs at that place.

(b) If the turn-so-far is so constructed as not to involve the use of a 'current speaker selects next' technique, then self-selection for next speakership may, but need not, be instituted; first starter acquires rights to a turn, and transfer occurs at that place.

(c) If the turn-so-far is so constructed as not to involve the use of a 'current speaker selects next' technique, then current speaker may, but need not continue, unless another self-selects.

(2) If, at the initial transition-relevance place of an initial turn-constructional unit, neither 1a or 1b has operated, and, following the provision of 1c, current speaker has not continued, then the rule-set a–c re-applies at the next transition-relevance place, and recursively at each next transition-relevance place, until transfer is effected. (Sacks et al., 1974, p. 704)

The rules are presented in order of occurrence, meaning that current speaker has the primary right to select the next speaker. Only when current speaker has not selected a next speaker do other participants get a chance to select themselves as speakers. This has the effect that speakers generally are only attributed one turn-constructional unit at a time, that is, they are allowed to produce one recognizably complete turn before speaker transfer can and usually should occur. Only if no other participants selects themselves to be the next speaker does current speaker get rights to continue.

Clearly this is not an exhaustive nor a deterministic description of turn taking in conversation. Speakers of a possibly complete turn can and do continue in violation of the *rules*, just as recipients will sometimes self-select in an environment where speaker-transition was not made relevant or where another participant has been selected as next speaker. Furthermore, speakers can be allowed to produce more than one turn-constructional component before transfer is possible and relevant, that is, the system can be temporarily suspended. But the system is treated as normative, that is, participants hold each other

accountable for adhering to it. At the same time they continuously re-establish it with every successful transfer of speakership.

With this system in place, we are also provided a "proof procedure for the analysis of turns" (Sacks et al., 1974, p. 728). When speakers produce an FPP, they select by conditional relevance a next speaker to provide a type-fitting response. Next-speakers will therefore be understood to be providing that type-fitting response. In other words, by providing a certain type of response, next-speakers displays their understanding of the type of adjacency pair that was initiated by the first-speaker and thus their understanding of the action produced by that first speaker. In fact each turn at talk is understood in relation to its prior, adjacent turn, unless it is designed as not to be so understood (Schegloff, 1988a). Producing an utterance subsequently to another utterance, that is, *next positioning* an utterance, is a primary means for making it understood as related to that prior utterance (Jefferson, 1978, fn. 8).

So it is in the next turn that participants reveal to each other how they understand one another, and it is there that we can find evidence for our analysis of the action that a turn is used to implement. This notion is central to the various analyses in this dissertation. In the next turn recipients display their understanding of a prior declarative yes/no-type initiating actions as for example a request for confirmation or an invitation to tell (see chapters 2 and 3); they distinguish between turns that are doing now-understanding and turns that are aimed at resolving knowledge discrepancies (see chapters 4 and 5); and they display their understanding of an answer as either informative or a proposal (see chapter 6). In all these cases the next turn thus provides evidence for our analysis of the action implemented in the prior.

1.3.2 Intersubjectivity in interaction

In the previous section I discussed the mechanics through which participants coordinate their actions. In this section I show that through these mechanics participants solve a problem that particularly sociology and philosophy have wrestled with for a long time: the problem of intersubjectivity. Simply put, the problem is as follows: Two or more participants need to coordinate their actions without being able to directly access each other's intentions and understandings: "[a recipient] knows merely that fragment of the [speaker's] action which has become manifest to him, namely, the performed act observed by him or the past phases of the still ongoing action" (Schutz, 1962, p. 24). This limitation clearly is central to any theory that has as its aims to provide an explanation of social interaction. As Schegloff (1992, p. 1296) explains: "without systematic

provision for a world known and held in common by some collectivity of persons, one has not a misunderstood world, but no conjoint reality at all." But no two individuals will ever have identical experiences or perspectives of anything, so how can two people rely on shared experiences or shared assumptions? We need a provision for a world held in common, when there can never be such a world.

Part of the explanation has to be sought in how participants in social interaction make themselves understood. They achieve this not only through language, but also rely on the context (consider again Schegloff's definition of the action-formation problem in section 1.1): any turn-at-talk will be designed and understood in relation to when, where, and by whom it is produced. Participants thus rely on what is often called the *common ground* they have with their co-participants (Stalnaker, 1978).

Understanding how participants build up and use the common ground is thus part and parcel to understanding action formation. In this section I discuss a prominent theory developed by Clark (1996) of how participants manage their common ground. Clark argues that because common ground is crucial for social interaction, an account of social interaction cannot rely on an intuitive appeal to the context. Instead we need a *proper theory* of common ground. While the theory Clark provides does allow for a more grounded analysis of action formation, I argue that it does not actually preclude an intuitive appeal to the context, and in fact that it still relies on commonsense assumptions about how participants manage their common ground.

Subsequently I discuss the procedural nature of intersubjectivity as it is applied and understood in CA (Schegloff, 1992). While there is clear overlap with Clark's approach as should become clear from the respective discussions, the focus in CA is not on how participants base their common ground in for example assumptions about communities and shared experience, but instead on how intersubjectivity is managed and grounded in the local sequential structure of the interaction.

Context and Common Ground

Clark (1996) is concerned with what participants in social interaction know and assume the other participants know and assume. Any action is designed for a specific participant or set of participants (Sacks et al., 1974, p. 727), and so speakers routinely make appeals to what they perceive as their common ground. Furthermore, interaction, as Clark understands it, is aimed at expanding the common ground; indeed, he argues that the size and shape of the common

ground of two participants reflects the intimacy of their relationship (Clark, 1996, p. 115): The more expansive the common ground, the more intimate the relationship. The question then is how is common ground brought about, and how is it managed in talk-in-interaction.

There are two fundamental points that Clark (1996) makes in his approach. He first provides a formal definition of common ground, which shows how common ground is established and managed in interaction. Subsequently he distinguishes between two types of bases on which participants make their assumptions about the common ground. I discuss them in the same order.

Common ground for Clark (1996) is a reflexive concept. This means that it is not enough that each participant has access to the same piece of information, but that they also know that each of them has access to that same piece of information. In addition, this reflexive knowledge requires a shared basis that indicates the same information to all participants; it is the assumed shared basis that justifies the assumption that some belief is part of the common ground. This has as an important implication that common ground need not be established through interaction. Two people can assume that given a certain shared basis, which invariably has to be assumed to be a shared basis, they have a shared belief and that shared belief is thus part of their common ground.

Consider the following situation. If my father and I are sitting at Wimbledon Center Court watching Federer play Nadal, we are presented with the same visual basis on which to make assumptions about what the other sees. So we can say that the belief that we are watching Federer play Nadal is part of our common ground.

But consider that the other 15,000 spectators have the same visual basis, and we would not want to argue that we have the same common ground, or a common ground at all, with all these other spectators. We merely share a basis based on which we could of course build a reflexive common ground. The difference is partly that my father and I are watching *together*; it is an activity in which we both participate and we are aware that this participation is shared. We are undoubtedly also aware of the rest of the crowd, but not as individual spectators. Our watching is therefore not a shared activity (see Sidnell, 2014).

Common ground is, however, not as simple as that. My father and I may be looking at the same thing, but that does not mean we can assume we see the same thing. I may see Federer dominating Nadal by playing the best tennis of his career, whereas my father may see an injured Nadal struggling to keep up. We are presented with the same visual basis, which serves as evidence both for our understanding that (a) we are watching Federer play Nadal and (b) we are watching Federer dominate Nadal or Nadal struggling respectively. But

while the visual basis may be strong evidence for (a) it can be relatively weak evidence for (b). So while we would probably say that (a) is almost certainly part of our common ground—tennis fans as we both are—we may be relatively uncertain about whether (b) is indeed part of our common ground.

The second aspect of Clark's discussion deals with how participants in talk-in-interaction come to a shared basis. He argues that common ground can have two types of bases: (i) the cultural community the participants belong to, what he calls "communal common ground"; and (ii) the direct personal experiences participants have had, what he calls "personal common ground" (Clark, 1996, p. 100ff.).

Community as a basis for common ground relies on the stratification humans make of society. We all belong to a vast set of different communities, and each one comes with assumptions about what other members of that community ought to know. In addition, we have knowledge of communities we do not belong to and assumptions about what people who do belong to those communities know. Similarly, we have assumptions about what people who do not belong to our communities would know of them. Based on the communities we and others belong to, we make assumptions about what they might know.

The personal common ground is of a different nature. It is based on the experiences that people share with one another: what people see and do together. It is the personal common ground that according to Clark defines the relationship between people. Two people who belong to the same communities do not need to be acquainted in any way. But the more they do together and learn about each other—that is, the more they increase their personal common ground—the closer they become.

Although Clark's formalization of the common ground seems a useful step, and the distinction between communal and personal seems a beneficial one, it is unclear how it achieves its goal: namely, to constrain our analyses. For any conversational contribution, Clark (1996, p.221) argues that participants work actively to ground it: 'to establish it as part of the common ground well enough for practical purposes.' But this does not mean that participants specify how they come to an understanding of an action. It merely means that for any utterance the recipient will have to provide positive evidence that it was adequately heard and understood. Depending on the type of contribution, the typical way of doing so is by simply providing a relevant next; completing the joint project. The successful completion of a joint project is the basis for adding that joint project to the common ground, but whatever assumptions the participants rely on while constructing their joint project is still under the surface of the interaction.

Clark takes issue with an undefined context, because then one basis for a mutual belief is as good as the other. With no formal constraints on the context, any explanation is mere speculation. Under Clark's proposal, we cannot simply appeal to the context, but we would have to point out some specific element in the context that participants use as the basis for their mutual beliefs: a common community or a shared experience for example.

And in fact in current CA work this is common practice: in discussions of data, researchers generally provide a minimal ethnography of the participants and the situation, inherently claiming that this is relevant for the participants' understanding of the interaction. But the relevance of this ethnography is not discovered by the researchers through some formal procedure. It is in fact based on a commonsense understanding of what in the context the participants orient to. While this analysis should subsequently be grounded in the participants' observable behavior, we can only make a reasonable appeal based on our own commonsense understandings of the interaction—unless of course they make explicit what aspect of the context they are appealing to.

For any turn-at-talk, the basis could be prior talk in the same conversation, it could be some prior shared experience, it could be communal knowledge, and so on. We cannot know on what basis participants make assumptions about their mutual beliefs. In fact, we don't know what the participants consider their common ground to be, beyond what they treat as shared in the interaction. The bases and reflexive understandings may be the mental representation of the common ground, but we have no way of verifying this, or deriving our analysis from it.

So for our analysis of the moment-by-moment understandings that are established through interaction, an intuitive notion seems as good as Clark's *proper theory*. Some specification is required, but that specification is still a matter of plausibility.

Procedural intersubjectivity

The previous section showed how Clark (1996) attempts to capture the bases that people use to ground their mutual beliefs on which they rely in interaction. But since interaction is required to build a common ground, it tells us nothing about how interaction itself is possible. We have what looks like a vicious circle: we ground our mutual beliefs through interaction, but we require at least some common ground, some mutual beliefs to be able to interact in the first place. Although Clark demonstrates how incorrect assumptions can be repaired as soon as they come to light, whereby we could revise the common ground, we

of course would then require the repair mechanism to be part of the common ground.

Speakers design their actions in a way that they can be understood by their recipient, and similarly recipients ascribe actions to utterances based on the assumption that that utterance was designed in a way that it could be understood by them. This requires intersubjectivity, and so understanding how intersubjectivity works is anterior (see Schegloff, 1992); its existence cannot simply be assumed if one is to understand how action formation and action ascription work:

> The question how a *scientific* interpretation of human action is possible can be resolved only if an adequate answer is first given to the question how man, in the natural attitude of daily life and common sense, can understand another's action at all. (Schutz, 1964, p. 20f.)

The view taken in CA can be traced back primarily to Schutz (1962) and Garfinkel (1952, 1967). Schutz treated intersubjectivity as a problem that is routinely solved in interaction by the participants assuming a "reciprocity of perspectives": (i) each has his or her own unique perspective, but those perspectives are interchangeable—person A's perspective would be the same as person B's if A were in B's position; and (ii) those differences in perspective are irrelevant until proven otherwise (Schutz, 1962, p. 11ff.). For Schutz, intersubjectivity is thus never guaranteed by some external factor like socialization in a common culture, but it has to be continuously assumed and negotiated (see Heritage, 1984b).[10] Garfinkel (1952, 1967) in turn built on these ideas, focusing on the importance of temporality that Schutz introduced in the study of intersubjectivity: "The appropriate image of a common understanding is (...) an operation rather than a common intersection of overlapping sets" (Garfinkel, 1967, p. 30).

The importance of this *procedural* nature of intersubjectivity was most clearly shown by Schegloff (1992) who argued that participants do not deal with *a* problem of intersubjectivity, but a recurrent *situated intersubjectivity*: "particular aspects of particular bits of conduct that compose the warp and weft of ordinary social life provide occasions and resources for understanding, which can also issue in problematic understandings" (Schegloff, 1992, p. 1299). As was discussed in section 1.3.1, the turn-taking system of interaction provides

---

[10]Seemingly independent of Schutz, Rommetveit (1974, p. 86) takes the same perspective when he states that "intersubjectivity has to be taken for granted in order to be achieved."

the participants with a proof procedure (Sacks et al., 1974): by producing a FPP speakers make conditionally relevant a type-fitting next action and the recipient's next utterance will be understood in light of this projection. Because recipients will display their understanding of the turn to which they addresses themselves—the action it implements, the social relationship it presupposes, its point of completion, and so forth—there is an opportunity for the speaker to address any perceived misunderstandings (Schegloff, 1992).

The *repair space* as Schegloff (1992; see also Schegloff, 2000) describes it provides for the following structure. At any transition-relevance place, the recipient of some turn (T1)—I will hereafter refer to the speaker of T1 as Speaker A and the recipient of T1 as Speaker B—has an opportunity to convey that he or she did not fully hear or understand that turn. By foregoing this opportunity, by not initiating repair, Speaker B tacitly conveys a belief that he or she understood A's turn. Furthermore, because of the adjacency relationship, B's subsequent response (T2) will display how B understood T1, thereby inherently providing A with evidence of how B understood T1. At the point where B's turn reaches possible completion, the system works in the same way. By not initiating repair, A tacitly conveys that he or she understood T2. And in the subsequent turn (T3) A will display an understanding of T2.

A now has evidence of how B understood T1 and B has evidence of how A understood T2. But B has no evidence that the understanding displayed in T2 of T1 is indeed adequate. But the system inherently provides for that. By not initiating repair participants tacitly convey that there is no repairable. Given that A has evidence of how B understood T2, there has been an opportunity for A to initiate repair had that understanding been somehow inadequate. So by not initiating repair A not only conveys that he or she adequately understood T2, but also that B displayed an adequate understanding of T1. In other words, by not initiating repair, both participants orient to a shared assumption of intersubjectivity: They treat their understanding as adequate and adequately shared (see Robinson, 2014).[11] The repair space can be schematically visualized as follows:

T1	A:	Q1		
T2	B:	A1	NTRI (*T*1)	
T3	A:	Q2	NTRI (*T*2)	Repair 3d (*T*1)

---

[11] A could after T2 also have explicitly ratified B's understanding by providing some sequence-closing third (Schegloff, 2007; see also Heritage, 2018; Houtkoop-Steenstra, 1985; Jefferson & Schenkein, 1977; Kevoe-Feldman & Robinson, 2012; Kevoe-Feldman, 2015; Koole, 2015; Schegloff, 1992; Tsui, 1989).

T4	B:	A2	NTRI (*T*3)	Repair 3d (*T*2)	Repair 4th (*T*1)
T5	A:	Q3	NTRI (*T*4)	Repair 3d (*T*3)	Repair 4th (*T*2)
T6	B:	A3	NTRI (*T*5)	Repair 3d (*T*4)	Repair 4th (*T*3, 1)

(Schegloff, 1992, p. 1327)

As this schema shows as long as repair is not initiated participants will continue under the assumption that they understand and are understood, that is, that intersubjectivity has been maintained. Only when repair is initiated is progressivity halted and do the participants have to work at re-establishing intersubjectivity.[12] The procedural approach to intersubjectivity saves the participants from the vicious circle of having to re-confirm that T1 was adequately understood, by confirming that T2 displayed an adequate understanding of T1, that T3 displayed an adequate understanding of T2 and that T2 thus displayed an adequate understanding of T1, etc. ad infinitum. People in their daily lives are not concerned with getting definitive proof; they look for evidence that is adequate for practical purposes:

> We may just take for granted that man can understand his fellowman and his actions and that he can communicate with others because he assumes they understand his action; also, that this mutual understanding has certain limits but is sufficient for many practical purposes.(Schutz, 1962, p. 16; see also Garfinkel, 1967)

Of course, such a method of bilateral assumptions is not fool proof, but it is remarkably efficient. Rarely do speakers initiate repair after next turn, that is, in third position. Repair in fourth position, or what is sometimes called post-sequence repair (Ekberg, 2012; Wong, 2000), is even more rare (see also chapter 4 in this dissertation). This may be in part because once the structurally provided for opportunities for repair have come and gone, there has to be a good reason to go back to fix a problem. Once a sequence has been successfully completed, the assumption of intersubjectivity has been interactionally validated. If at some later point one of the participants realizes that there was a misunderstanding in some earlier sequence, fixing it would mean halting the progressivity of an ongoing, possibly completely unrelated activity (see Stivers & Robinson, 2006

---

[12]Of course, they still rely on the same mechanism of repair. But participants proceed under the assumption that this is indeed the case, and so some level of intersubjectivity is maintained. A true and complete breakdown of intersubjectivity, if such a thing exists, can inherently never be repaired. It would require that some or all of the participant are not even aware of the other as a person attempting to engage in coordinated action.

on the preference for progressivity in interaction). Seeing as the sequence came off unproblematically even with the misunderstanding, there is no "need" to initiate repair.[13] The other side of the story is that most problems are simply resolved by the point that a slot for repair after next turn, let alone fourth position repair, comes along (Schegloff, 1992, 2000).

This discussion shows that repair after next turn is indeed as Schegloff (1992) says in the title of his article "the last structurally provided defense of intersubjectivity in conversation" and that intersubjectivity is procedural. By recognizing that intersubjectivity is procedural in nature, it should be clear that we cannot use notions such as the "literal meaning" of an utterance as a basis for describing how participants make their actions understood and accountable. Such a concept presupposes an invariant and objective meaning of an utterance that will inherently be shared by fluent speakers of a language; it puts the onus of intersubjectivity back on socialization in a common culture. Consider instead that any turn-at-talk is produced in a larger sequence of actions and is therefore inherently "context-shaped": Participants understand their interlocutors' turns-at-talk and design their own so as to be understood in relation to not only the immediate prior turn, but the larger sequential structure in which those turns are embedded (Heritage, 1984b, p. 242). Both the process of action formation and that of action ascription thus rest on the reciprocal assumption that the action as it is formed by a particular speaker will be understood by its orientation to the recipient to whom it is addressed (Sacks et al., 1974, p. 727).

1.3.3 Interactional Linguistics

In the previous sections I have focused how CA approaches the organization of interaction. But so far I have not discussed how this pertains to language and linguistic structure. Although CA is concerned with the practices participants use to make their actions in talk-in-interaction recognizable and accountable (Levinson, 2013; Mazeland, 2013; Schegloff, 2007; Sidnell, 2013), language was initially not a topic of study in and of itself (Fox, Thompson, Ford, & Couper-Kuhlen, 2013). CA belonged first and foremost to the field of sociology, and the study of language was limited to linguistics. But it should be obvious that we cannot have one without the other; that is, language is one of the, if not the central tool with which participants communicate. To understand talk-in-interaction we cannot but study language.

---

[13]Of course, what is considered needed is up to the interactants, and talk is not organized by orientation to some formal logical rules and procedures.

Less obvious may be that language and interaction constitute a two-way street. Linguistics has since the Chomskyan revolution often been thought of as modular, with the study of linguistic structure and linguistic meaning—syntax, semantics, phonetics, and so forth—being wholly distinct from the study of language use—pragmatics. Going back to Humbholdt and de Saussure, Chomsky (1965) argued that we need to distinguish between *competence*, what a speaker knows of the language, and *performance*, the actual use of the language. And for Chomsky this was a one way street: competence was needed for performance, but was not influenced by it. That is, we learn language through interaction, but that interaction does not affect the structure of the language. Performance should in fact be ignored, since it is influenced by more than just competence and is open to such nuisances as repair:

> Linguistic theory is concerned primarily with an ideal speaker-listener, in a completely homogeneous speech-community, who knows its language perfectly and is unaffected by such grammatically irrelevant conditions as memory limitations, distractions, shifts of attention and interests, and error (random or characteristic) in applying his knowledge of the language in actual performance. (....)
>
> Observed use of language or hypothesized dispositions to respond, habits, and so on, may provide evidence as to the nature of this mental reality, but surely cannot constitute the actual subject matter of linguistics, if this is to be a serious discipline. (Chomsky, 1965, p. 3f.)

But if we are to take the importance of temporality and reflexivity in interaction seriously, such an approach to linguistic structure is fundamentally flawed (Auer, 2009; Hopper, 1988). It assumes that linguistic structure is largely invariant, and that once fully acquired it is fixed in the mind of the speaker. Language is in this view distinct from the world, a completely independent object that can be investigated in isolation of extra-linguistic factors. But language is part and parcel to social interaction, indeed, conversation is the natural home of language (Sacks et al., 1974): Language shapes and is shaped by conversation (Couper-Kuhlen & Selting, 2001). The structuralist view of language as a set of forms independent from the real world is as Linell (2005) puts it the result of a Written Language Bias (see also Rommetveit, 1988).

That is not to deny that speakers design their utterances through structural means; for an action to be recognizable, its design should be understandable

by the recipient. Speakers within a community will thus inherently use the same linguistic practices to generate their actions. As Sacks (1995, p. 226) argued: "A culture is an apparatus for generating recognizable actions; if the same procedures are used for generating as for detecting, that is perhaps as simple a solution to the problem of recognizability as is formulatable." But that does not mean that speakers in a community have a uniform, mental grammar. Instead they use their prior linguistic experiences to generate new utterances: "the collective sum of actual speakers' experiences (...) is (...) the basis for the creation of new utterances without determining their structures" (Auer & Pfänder, 2011, p. 4). The result is that speakers rely on what could be called a cultural grammar.[14] But this grammar is a communal grammar, meaning that it is not in the mind of the speaker, but that it is continuously being reshaped and reconfirmed by participants in interaction. Grammar is therefore never finished, but always emergent (Hopper, 1987, 2011, 2012).

The study of linguistic structure in interaction has come to be known as Interactional Linguistics (Couper-Kuhlen & Selting, 2001, see Fox et al., 2013 for an overview). Although IL is strongly associated with CA, and the points of interest and study often overlap, the two can be considered distinct. CA focuses on the social organization of interaction, whereas IL is aimed primarily at furthering our understanding of language and linguistic structure by studying how they emerge and are used in social interaction (Ford, 2010).

Consider for example the turn-taking system for conversation (Sacks et al., 1974). It provides for a conversation with limited overlap and silence, but to do so successfully participants need to be able to project when a turn will reach possible completion. Linguistic structures such as syntax and prosody are a partial solution to that problem (Ford & Thompson, 1996; Fox, 2001; Huiskes, 2010; Schegloff, 1996c; Selting, 2000; Steensig, 2001; Tanaka, 1999): by producing turns in a consistent structural manner, the end of a turn becomes projectable, which allows for smooth turn transition.[15] Additional proof for such an analysis is found in cases where participants break with the normal projectability that language provides. For example, when speakers move to produce more than one turn-constructional unit in an environment where they are only granted room for one, various linguistic tools are used to annul the

---

[14]It stands to reason that humans from different cultures confronted with the same interactional problems, will devise similar solutions, resulting in the appearance of a universal grammar.

[15]In addition to form, one needs to consider action. That is, a turn needs to reach syntactic, prosodic, and pragmatic completion, and these aspects are considered together as a sort of gestalt (Ford & Thompson, 1996; Selting, 2000; T. Walker, 2017).

normal, projectable transition point (Local & Kelly, 1986; Local & Walker, 2004, 2012; Schegloff, 1982; G. Walker, 2017a).

But linguistic structure is not used just to manage turn taking. Much of the work in IL investigates how actions are constructed in specific sequential positions; the action-formation problem is one of its central points of research. See for example Fox and Heinemann (2016) on how lexico-syntactic and sequential aspects can be considered together in the doing of requests; Benjamin and Walker (2013) on how recipients use high-rise fall repetitions to claim that the prior turn is in need of correction; or Couper-Kuhlen (2014) on how grammar provides cues for whether an action under construction is a proposal, request, offer, or suggestion.

In this line of inquiry it is important to consider that linguistic structures are not simply retrieved from a mental grammar and implemented in interaction under some series of pragmatic constraints, but that actions can be said to have their own positionally-sensitive grammar (Schegloff, 1996c). This notion is crucial to the analyses presented in this dissertation. Although I deal with what may seem to be readily given syntactic units—primarily declarative and interrogative sentence types—these are not considered invariant units generated by a mental grammar, nor do they come with a fixed meaning independent of their environment of use. Their design and the actions they implement are adapted to a specific sequential environment (see Deppermann, 2011a): they are for example responsive to an informing (see particularly chapters 2 and 5), or follow closure of some other activity (see chapter 3). Furthermore, my analyses reconfirm that utterances are designed to deal with the local exigencies of the interaction (see particularly chapter 6; see also Mazeland, 2013).

Treating language as an isolated system thus inherently leads to an inadequate understanding of both language and the interactional organization in which it is used, because language is never isolated from its use. Hopper (2011, p. 32f.) drives this point home by comparing language to jazz music, arguing that musicians rely on themes they learned through training or listening to other music. This metaphor is very apt indeed. For any form of musical improvisation—not just jazz, but also for example blues—musicians rely not just on the music they have heard before, but also on their knowledge of musical scales in a certain key, and what are called *licks*, fixed—what might be called grammaticalized—series of notes.

Just as musicians do not need an a priori musical grammar to generate new and unique melodies, but can instead rely on prior experience and a limited set of constructions, so too do speakers not need an a priori linguistic grammar to produce language. And just as musicians adapts their improvised music to

that of their fellow musicians, both what they play and the rhythm at which they play it, so too do speakers continuously adapt their speech production to the local context, both their interlocutors and the sequential environment. Like music, language is the product of creativity in using recognizable patterns.

Discussion

In these sections I discussed the methods of Conversation Analysis and Interactional Linguistics, and showed how they are used by participants to achieve an understanding in and of social organization, linguistic structure, and most importantly for this dissertation, action formation. Following Schutz (1962) and Schegloff (1992), I argued that intersubjectivity—that is, a world held in common—is anterior to these problems, but that it is provided for by the procedures that participants use in talk-in-interaction. Similarly, linguistic structure is not a priori given, but emerges in and through interaction (Hopper, 1987, 1988, 2012)—it shapes and is shaped by conversation (Couper-Kuhlen & Selting, 2001)—in order to deal with the various interactional problems that participants need to address in order to coordinate their actions and make talk-in-interaction possible (Mazeland, 2013).

1.4   Previous work

Researchers from various fields have long recognized that interrogative syntax is neither necessary nor adequate to implement requests for information; declarative questions have long been known to be part of the interactional repertoire of Western languages like English, Dutch, and German.[16] Given the initial assumption that syntax was supposed to have a literal force, the explanation for how these declarative questions came to be made recognizable and understood as questions was sought in their prosody: they were thought to have a rising boundary pitch. But this assumption too has been frequently shown to be inadequate (Geluykens, 1987, 1988; Beun, 1989b; Couper-Kuhlen, 2012; Seuren, Huiskes, & Koole, 2015; Strömbergsson, Edlund, & House, 2012). So if the grammatical features of a turn cannot account for its questioning status, how then can they be so understood?

In the following sections I discuss four approaches to this problem: one based in Speech Act Theory, two in Formal Semantics, and one in Conversation

---

[16]I use *Western* not as a synonym for *Western Germanic* but to reflect the initial focus of linguists and language philosophers on the languages of their own Western cultures.

Analysis. Note that while all four approach the problem from a different angle, they deal with roughly the same problem. It should thus perhaps come as no surprise that while the methods differ, some of the answers they provide do not. In addition to discussing how the findings of these approaches relate to the analyses presented in this dissertation, I discuss the implications of these similarities, suggesting that it could point to a possible reconciliation of the various methods.

1.4.1   Preferred interpretations

In his dissertation titled *The Recognition of Declarative Questions in Information Dialogues* Beun (1989b, p. 1f.) presents a series of studies that were aimed at understanding "how listeners in natural dialogues identify the question function of a DQ [Declarative Question] and which information is conveyed if a declarative form is used instead of an interrogative." His focus is thus both on action formation, what a speaker does by using a declarative, and action ascription, how a recipient comes to understand a declarative as a question. Although his framework is Speech Act Theory, his studies are not just philosophical or theoretical, but largely experimental, and rely on recordings of actual, albeit not completely naturally occurring, interaction.

The corpus used by Beun consists of a series of recordings of telephone conversations between the Schiphol Information Desk and people seeking information on such issues as plane arrival or departure times, flight numbers, or traveling options to the airport (Beun, 1985). The recordings are, however, experimental in nature: the callers are not actual service seekers, but participants in a study. The service provider is trained in the job though, making the interaction at least partly natural. From these conversations, Beun selected all Declarative Questions according to the following definition (Beun, 1989b, p. 23f.):

(3)   An Utterance U is a declarative question if:
   a. The sentence type of the sentence expressed by U is declarative (or if the sentence is elliptical the sentence type is at least non-interrogative and non-imperative).
   b. The utterance U, uttered by S [Speaker], is about a topic on which S believes that H [Hearer] is the expert.
   c. S believes that S and H mutually believe that H is the expert on the topic.

From the Declarative Questions that were collected according to this definition, cases that had what were considered clear question markers were removed: cases for example with a rising boundary pitch; cases with turn-final particles such as *hè*, and *toch* (see Enfield, Brown, & de Ruiter, 2012; Foolen, 1994); and cases where the speaker in the design of the question conveyed who was considered expert through formulations such as *you said*. The final corpus was used in a series of experiments, aimed at finding out what it was in their turn design that made these utterances understood as questions.

Findings

In the first experiment (see also Beun, 1989a) the questions were cut from the recording and presented to participants who had to decide whether it was a question or not, or whether it was an answer or not. This experiment suggested that various aspects of turn design, such as specific particles or self-repair, could help make an utterance understood as a question. These features were removed from the utterances, and used in a second study where participants again had to decide whether or not the utterance was a question, or whether or not it was an answer. The findings suggest that conjunctions like *en* ('and') and *dus* ('so'), as well as turn-initial *oh* help make an utterance understandable as a question and not an answer.

By removing the potential question features that were uncovered in the first two experiments, certain aspects of the prosody were inherently also cut out. In order to compensate for this shortcoming, a follow-up study was done in which the questions were presented in written form (see also Beun, 1990b). In this case particles like *toch* were included. The findings were again the same: *en*, *dus*, and *oh*, as well as *toch* significantly contributed to making an utterance understood by the participants as a question.

The explanation Beun gives is two-fold. Particles like *en* and *dus* conjoin two utterances, but the participants knew that these particles had to be turn-initial. That is, they were not used by one speaker to link two parts of one turn together, for example two parts of an answer, but they were used to show how the turn relates to the prior talk. A particle like *dus* formulates an inference from the prior talk by the other speaker, and so according to Beun indicates that the interlocutor is the expert on the topic, and hence that the turn is a question, not an answer (see condition (b) of his definition). Particles like *oh* and *toch* on the other hand signal surprise or conflicting beliefs, and so turns that contain these particles are understood as requests for clarification.[17]

---

[17]Chapter 4 of this dissertation presents a study of *oh*-prefaced declaratives, showing that

In addition to investigating the linguistic factors that contributed to participants' understanding of declarative questions as questions, Beun studied how context could contribute (see also Beun, 1990a). In order to explore this issue, he selected eighteen dialogues of which he presented participants with two written versions: the original and one with a slightly modified sequential context. Participants then had to say whether they preferred a declarative or interrogative question given the context, and how certain they thought the speakers who asked the questions were. While Beun found a strong relation between perceived knowledgeability and syntactic format, this was not one-to-one. He found that declaratives are preferred when the questioned information is given in the context, but that with for example negative interrogatives this no longer applies: a negative declarative is preferred over a negative interrogative even if the speaker is not perceived to have strong beliefs on the addressed issue.

Based on these experimental findings, Beun proposes (see also Beun, 1994) that the function of an utterance as either an assertion or a question depends on what he calls the *function structure* which consists of various turn design features of the utterance: sentence type, particles, and prosody.[18] According to Beun, the function structure is a function that applies to propositions and which generates a communicative act. Any combination of feature values will generate a function structure with a preferred interpretation. If on the basis of the context this preferred interpretation is ruled out, the recipient will understand the utterance to be doing the less preferred action. For example, if a speaker produces an utterance with declarative syntax and no linguistic question features, the preferred interpretation will be that of an assertion. If, however, it can be proved in the context that the speaker does not intend to let the hearer know something, the utterance will be understood as a question.[19]

Discussion

Beun's approach to the puzzle of declarative questions is interesting from an interactional perspective, because he works with recordings of actual talk-in-interaction. These may be experimental, but as recent discussions in CA

---

they are indeed used as questions, in the sense that they make confirmation relevant. However, I argue that they are used to do now-understanding.

[18]For prosody he only considers the boundary pitch of an utterance (see G. Walker, 2017a).

[19]The list of interpretations is not limited to two possibilities. If an utterance can subsequently be shown to not be a question, the recipient will move on to the following possible interpretation on the list.

have suggested, that does not mean the data is unnatural (de Ruiter & Albert, 2017; Kendrick, 2017, see also Stivers, 2015). Whereas Schegloff (1988b) has argued that SAT is inherently irreconcilable with CA, because SAT takes single, context-free utterances as its unit of analysis, Beun also considers the context, and actually makes it a specific issue of investigation. His findings about the role of turn-initial lexical and non-lexical items such as *dus*, *en* and *oh* in the design Declarative Questions have indeed been confirmed in recent work (Seuren et al., 2015).

Having people from outside the interaction judge the interactional function is also not necessarily problematic. Beun does not offer the findings based on their action ascriptions as evidence of what an action is doing for the participants in the interaction, but as an initial step to explore what might be the relevant linguistic factors for making an utterance into a question, and what is achieved with those questions. There are however some important criticisms to make.

One crucial problem is that participants in the experiments had to choose whether an utterance was or was not a question, and whether an utterance was or was not answer. This does ignore the sequential import for participants' understanding of interaction (Schegloff, 1988a, 1988b, 1996b, see also Sacks et al., 1974; Schegloff & Sacks, 1973). When confronted with an utterance, any recipient will be aware of what came before that utterance. If the recipient has just asked a question in the turn before the speaker's turn, the issue of whether the speaker's turn is a question or an answer will not arise. It will be understood as responding to that question.[20] The manner by which the participants in the experiment came to their action ascription is thus fundamentally different from how participants in talk-in-interaction come to their action ascriptions.

It is also hard to see how his conclusions can be reconciled with CA. Beun (1989a, p. 125) suggests that CA and SAT are not competing approaches but should be supplementary. But the way he formalizes action formation and ascription is fundamentally different from how CA views this problem. For one, it presupposes that an utterance can only be used to do one action, but utterances can do a multitude of things (Enfield, 2013; Levinson, 2013; Schegloff, 1988b; Sidnell & Enfield, 2014; Sidnell, 2017a). Second, the broad range of linguistic and embodied practices that people use to design their actions would mean there is an exponentially larger combination of feature values and thus function structures for which speakers would need to remember their

---

[20]There is always the option that the speaker is initiating repair on the recipient's preceding turn (Benjamin, 2013; Dingemanse et al., 2015; Schegloff, Jefferson, & Sacks, 1977), but that is a very different distinction than the one proposed by Beun.

preferred interpretation as well as their less preferred interpretations and their respective orders. Third, as Schegloff (2007) points out in his definition of the action-formation problem, the size of the class of actions is unknown, and seeing that participants can perform such actions as *confirming allusions* (Schegloff, 1996a) we can assume it is a pretty large set.

That is not to say that Beun's approach is completely distinct from anything we find in CA. Recent studies have suggested that recipients use the various practices with which speakers design their utterances to infer what type of response is being projected (e.g., Sidnell & Enfield, 2014; Sidnell, 2017a, see also chapter 5 of this dissertation). So instead of going through a list of preferred interpretations, recipients can immediately infer what a speaker is doing; Enfield (2013) calls such a combination of practices that implements a very specific action a *praction*. The list of possible interpretations is also very limited, because in a certain sequential context, there is only a limited number of moves that speakers can make that would be considered coherent; their utterance will be understood in light of the prior talk (Schegloff & Sacks, 1973; Sacks et al., 1974).

The problem it seems is the assumption underlying the methodology. By viewing language through the lens of SAT, Beun treats action ascription as a matter of ascribing mental states—intentions—and attempts to capture that process in a formal system. As he says himself, he proposes "an extended version of the literal force hypothesis" (Beun, 1989b, p. 118), but only for cases where the context supports that literal force—that is, the preferred interpretation. But as I discussed in section 1.2, the LFH is not well-suited to deal with actual talk-in-interaction (Levinson, 1983; Schegloff, 1988b). Formalization is a logical approach if one wants to teach a computer how to interact with people, but it does not consider how human participants achieve a local understanding of their actions. It is thus hard to see how, other than his observations about turn design, Beun's findings could be integrated into a CA method.

1.4.2   Semantics of sentence types

In line with Gazdar's (1981) idea of dropping the LFH, we find a series of approaches in formal semantics that investigate the relation between linguistic form, semantic meaning, and pragmatic function. A few of these approaches deal specifically with the distinction between declaratives and interrogatives, as well as action formation and ascription, albeit indirectly: Gunlogson (2001, 2008) proposes an analysis that relies on a commitment-based discourse, and Farkas and Roelofsen (2017) propose an analysis grounded in Inquisitive Se-

mantics (Ciardelli et al., 2013).

Although the proposal by Farkas and Roelofsen (2017) has broader descriptive and explanatory adequacy—it accounts for a broader range of phenomena and generally gives more accurate predictions (see Chomsky, 1965)—I will discuss both approaches. Partly because Farkas and Roelofsen (2017) rely on the commitment-based discourse proposed by Gunlogson (2001), but also because in a few cases Farkas and Roelofsen's proposal makes predictions that Gunlogson correctly rules out.

While these approaches may strike conversation analysts as deeply flawed for their methodological assumptions (see the discussion in section 1.4.2 below), I discuss these theories for two reasons. First, they address research questions very similar to the ones addressed in this dissertation, and in a way they could be said to simplify the issue of action formation and ascription: a semantic interpretation should facilitate any pragmatic interpretation. Second, discussing these theories allows for a comparison with a CA approach that, as should become clear in section 1.4.3, has surprising similarities considering these methodological differences. I argue in fact that the ideas discussed here could serve to further develop an interactional grammar.

I first discuss the proposals by Gunlogson (2001, 2008) and by Farkas and Roelofsen (2017), and then discuss how they fare when applied to conversational data. In closing I argue that these approaches could be seen as possible grammars in the spirit of Schegloff's (1996c) proposal for multiple positionally-sensitive grammars.

Commitment-based discourse

In her dissertation, Gunlogson (2001) addresses the following, by now familiar, issue. Given that the same declarative sentence can be used both to make assertions and ask questions, what semantics do these sentences need if we cannot assume a shared semantics that is simply resolved differently in the context? That is, what are the conventional discourse effects, or what is the force (Frege, 1918/1956), that these sentences have? Gunlogson introduces this problem using the following minimal pair, (4) having a final rising pitch, and (5) having a final falling:

(4)   It is raining?

(5)   It is raining.

Gunlogson states that while (4) is the more natural question due to its rising boundary pitch, pitch alone cannot account for its questioning function, since (5) can also be used to question. An analysis of this minimal pair should therefore account for why (4) is the more natural question, without excluding (5) as a potential question. The conventional discourse effects of the respective sentences thus cannot include pragmatics, but should make certain pragmatic interpretations more straightforward.

The import of Gunlogson's analysis is thus primarily for the formation side of the action-formation problem that this dissertation is concerned with: what the circumstances are in which speakers (can) use declarative syntax with either rising or falling boundary pitch to make relevant confirmation, and how such a choice affects the discourse context. But inherently she also deals with action ascription: how recipients respond if one format is chosen instead of another in a specific context.

Gunlogson's account for how these two sentences differ relies on her approach to the discourse context of the interaction. Following Stalnaker (1978), Gunlogson (2001, p. 43) takes the discourse context to be the ordered pair of a set of possible worlds in which the beliefs of the participants are true—the commitment sets (cs) of the participants:

(6)   Let a discourse context $C_{\{A,B\}}$ be $< cs_A, cs_B >$ where:
      A and B are discourse participants
      a. cs_A of $C_{\{A,B\}}$ = {w ∈ W: the propositions representing A's public beliefs are all true of w}
      b. cs_B of $C_{\{A,B\}}$ = {w ∈ W: the propositions representing B's public beliefs are all true of w}

While the definition makes reference to public beliefs, the discourse context is not limited to the propositional content of contributions made in the interaction. Gunlogson assumes a context in which participants rely on a far broader set mutual background assumptions, very similar to how Clark (1996) defines the common ground (see section 1.3.2), but that is not pertinent to her discussion.

The solution Gunlogson (2001, p. 36) derives from this definition is, however, somewhat unsatisfactory. She argues that "rising declaratives commit the Addressee to the proposition expressed, whereas falling declaratives commit the Speaker." This means that by producing a sentence like in (4), the Speaker

adds the propositional content, $\mathcal{P}$, of (5) to cs_{Addr}. By producing a sentence like (5) on the other hand, the Speakers adds its propositional content to cs_{Spkr} (Gunlogson, 2001, p. 52).

Clearly an issue is how a speaker can commit an addressee to some proposition, and it is an issue Gunlogson (2001) leaves unresolved. In a more recent paper, Gunlogson (2008) takes a significantly different approach, while limiting her scope to what she calls Initiating Declarative Questions, that is, declarative questions that are not directly relating to prior talk. She still relies on the notion of commitment sets (see Hamblin, 1971), but she adds the concept of *source sets* (ss). A participant can, but need not be, a source for a proposition. This means that for any proposition added to a commitment set, a speaker may also add that proposition to his or her source set.

This extension of the model allows for a very simple distinction between the prototypical statement[21] and the prototypical question. Statements are normally done with declaratives; with a declarative the speakers adds the proposition φ to both his or her commitment set, cs_{Spkr}, and source set, ss_{Spkr}. Questions on the other hand are normally done with polar interrogatives; with a polar interrogative the speaker makes no changes to either set, but signals a dependency with respect to either φ or $\neg\varphi$. After an answer has been provided, the addressee adds either φ or $\neg\varphi$ to both cs_{Addr} and ss_{Addr}, and by acknowledging the answer the speaker will thus add either φ or $\neg\varphi$ to cs_{Spkr} but not to ss_{Spkr}.

This definition obviously leads to a new problem: there is no inherent difference between declaratives that are used as statements and declaratives that are used as questions. In both cases the speakers adds φ to both cs_{Spkr} and ss_{Spkr}. In order to resolve this Gunlogson (2008, p. 125) introduces the notions of *implicit source* and *implicit authority*:

(7)   An agent α is an *implicit source* for φ iff:
    a. α is not committed to φ; and
    b. It is inferable in the discourse context that if α commits to φ, α will be a source for φ.

(8)   An agent α is *implicitly authoritative* with respect to φ iff α is an implicit source for both φ and $\neg\varphi$.

This definition means that when neither participant is yet committed to either φ or $\neg\varphi$, but it is clear from the context which participant is going to be

---

[21]Gunlogson (2008, fn. 15) uses *statement* instead of *assertion*, as the latter is frequently used in semantics to refer to joint commitments, not individual commitments.

or would be the source for φ or $\neg\varphi$, that person—here called α—is implicitly authoritative.

Note that this definition does not yet distinguish between declarative statements and declarative questions; in fact it does not even distinguish between declaratives and interrogatives: Using a polar interrogative treats the addressee as implicitly authoritative. For declarative questions specifically, we need an additional criterion: *contingent commitment* (Gunlogson, 2008, p. 128). A declarative will be understood to be doing questioning if the speaker's commitment is contingent on subsequent confirmation by the more authoritative addressee:

(9)   A discourse move μ committing an agent α to φ is *contingent* upon
       ratification by an agent β, $\alpha \neq \beta$, if:
       a. β is implicitly authoritative with respect to φ at the time of μ
       b. It is inferable in the discourse context that $\alpha's$ commitment to φ will
           be withdrawn unless the discourse move immediately succeeding μ
           has the effect of committing β to φ as a source

An interesting effect of this definition is that rising and falling declaratives no longer differ from one another: they have the same semantic interpretation and the same semantic content. In her dissertation Gunlogson (2001) aimed for a semantics that treated rising declaratives as more similar to polar interrogatives than to falling declaratives. Now the semantics of rising and falling declaratives are the same, whereas that of polar interrogatives is very different. That is not to say that rising and falling declaratives are identical. There are discourse contexts where a falling declarative would be infelicitous while a rising would be acceptable; falling declaratives are only allowed in a subset of the contexts that allow rising declaratives, and those can in turn be used only in a subset of the contexts that allow polar interrogatives.

Rising intonation in this proposal explicitly marks the utterance as contingent, meaning that by using a rise the speaker conveys to the addressee that his or her commitment to φ depends on subsequent ratification of that commitment. Pitch only has a facilitating role, not a semantic one. The function of a declarative as doing questioning thus depends on the participants' understanding of the context. A falling declarative can only be used as a question when it is clear that the recipient has implicit authority, a rising declarative on the other hand requires less strict conditions, because it facilitates the understanding of the action as a contingent one.

To sum up: Gunlogson (2001, 2008) proposes that with declaratives the speakers adds the proposition to his commitment set and his source set. De-

pending on the context, either the speaker or the addressee will be mutually understood to have implicit authority on the issue raised. If the speaker has authority, the declarative will be understood as a statement. If on the other hand the addressee has implicit authority and the speaker's commitment is understood to be dependent on the addressee's ratification, the declarative will be understood as a question. A rising intonation facilitates this understanding, which means that rising declaratives can be used as questions in more ambiguous contexts in which it need not be clear who has implicit authority.

Inquisitive Semantics

Farkas and Roelofsen (2017) are like Gunlogson (2001, 2008) interested in how different grammatical formats can be used to ask questions. The focus of their analysis is thus also on action formation; action ascription is dealt with only tangentially. Farkas and Roelofsen (2017) take sentence types to consist of both syntax and boundary pitch. The explanatory scope of their proposal is however far broader than that of Gunlogson (2001, 2008): The analysis is not focused on one specific grammatical format, but on declaratives, polar interrogatives, and tag interrogatives, and each with rising and falling intonation. An additional distinction is that where Gunlogson (2008) requires each sentence type to have a different convention of use, Farkas and Roelofsen (2017) propose that all sentence types have the same basic convention of use, but can have an additional convention of use depending on the specific sentence type: some forms are more optimal than other in their *economy of form* and the degree to which they insure *communicative success*. I provide a somewhat simplified overview of their proposal in this section.

Farkas and Roelofsen (2017) focus on two levels of analysis: (i) the semantic interpretation, which takes the meaning and composition of the words of an utterance and gives the semantic content; and (ii) the convention of use, which takes the semantic content of an utterance and gives its conventional discourse effects. While they do also discuss some pragmatic discourse effects, those are not the focus of the paper.

For the semantic interpretation of an utterance, Farkas and Roelofsen (2017, p. 257ff.) assume that there are two kinds of *clause type markers*: DEC/INT, which is given by the syntax of a clause; and CLOSED/OPEN, which is given by the boundary pitch, CLOSED being final falling pitch and OPEN final rising. A clause that is either INT, OPEN, or both is inquisitive, whereas a clause that is both DEC and CLOSED is informative. This means that only falling declaratives have an informative semantics, while rising declaratives, tag interrogatives, and

polar interrogatives have an inquisitive semantics.

This does not mean that all inquisitive sentence types are equivalent: some are optimal, or unmarked, whereas others are marked. Since there is only one type of informative sentence type, the falling declarative, it is inherently an optimal form. For inquisitive sentence types, however, there are five forms: rising and falling polar interrogatives, rising and falling tag interrogatives, and rising declaratives. Given the condition of *economy of form*, tag interrogatives are more marked than polar interrogatives, because they have a more complex form. Rising declaratives are also more marked than polar interrogatives, because the latter are better at insuring *communicative success*: they provide the addressee with two cues that they are inquisitive, while rising declaratives have only one.[22] This results in the following classification (Farkas & Roelofsen, 2017, p. 265):

(10)   *Markedness classification*
           a. Optimal, unmarked forms:
                  • falling declaratives
                  • polar interrogatives
           b. Marked forms:
                  • rising declaratives
                  • tag interrogatives

For the discourse effects of an utterance, Farkas and Roelofsen (2017, p. 265ff.) propose that there is one basic convention of use that applies to all sentence types. But marked sentence types get an additional, special effect that is connected to the specific sentence type used. For this special effect, they introduce the concept of *evidence*, whose value depends on the sentence type: [zero, low, moderate, high].

The basic conventional discourse effect is taken from inquisitive semantics (Ciardelli et al., 2013). Unlike traditional semantic approaches that follow Stalnaker (1978), inquisitive semantics treat a proposition $\mathcal{P}$ of a sentence φ not as a set of possible worlds, but as a set of *information states* that consist of an informative and inquisitive content that support the sentence and that are modeled as a set of possible worlds. By producing a sentence φ, a speaker steers

---

[22]Since falling polar interrogatives and rising declaratives seem equivalent in this sense—both provide only one cue—one would expect that both are more marked than rising polar interrogatives. Farkas and Roelofsen (2017) skip over this issue: They provide a clear argument why rising polar interrogatives are less marked than rising declaratives, but say nothing on rising declaratives vis-a-vis falling polar interrogatives. Most likely syntax is assumed to take priority over prosody.

the conversation towards a context in which the expressed state s is supported, that is, where the participants agree that the actual world, w_a, is a member of s. Secondly, the speaker commits him-/herself to the claim that w_a is a member of some element in $[\![\varphi]\!]$, the proposition expressed by φ (Farkas & Roelofsen, 2017, p. 265f.):

(11) **Basic convention of use**
      If a discourse participant x utters a declarative or interrogative sentence φ, the discourse context is affected as follows:
      (a) The proposition expressed by φ, $[\![\varphi]\!]$, is added to the **table**
      (b) The informative content of φ, $\bigcup [\![\varphi]\!]$, is added to **commitments(x)**.

By uttering a falling declarative sentence the speaker adds α to his or her commitment set and puts $\{\alpha\}^{\downarrow}$, the proposition expressed by the sentence, on the table, thereby steering the conversation to a context where both participants mutually agree that w_a is in α. Since the speaker is already committed to this state, the recipient simply needs to provide some form of acknowledgment, thereby also adding α to his or her commitment set.

By uttering a polar interrogative the speaker also commits to the informative content of the proposition expressed by that utterance. But as the sentence is a polar interrogative, that proposition takes a different form: $\{\alpha, \bar{\alpha}\}^{\downarrow}$. The informative content in this case is equal to all possible worlds, W, so the speaker simply commits to a state in which w_a is part of W, which is trivially informative. The speaker also puts the proposition expressed by that utterance on the table, steering the conversation toward a context where both participants agree that w_a is either part of α or $\bar{\alpha}$. The speaker is thus neutral with respect to which is true, and the addressee has to provide a response that helps settle this issue.

The marked sentence types have the same basic convention of use. This means that the marked inquisitive sentence types all express the proposition $\{\alpha, \bar{\alpha}\}^{\downarrow}$. By producing an utterance that has either of those marked sentence types the speaker thus steers the conversation toward a context where w_a is part of either α or $\bar{\alpha}$ and conveys the trivial information that w_a is contained in W.

In addition, there is a special effect whereby the speaker adds the *evidence possibility* for α to his or her evidence list **evidence(x)**. For a rising declarative, this takes the form of $\langle \alpha, [\text{zero, low}] \rangle$, for a rising tag interrogative it is $\langle \alpha, [\text{low, moderate}] \rangle$ and for a falling tag interrogative it is $\langle \alpha, [\text{high}] \rangle$. Simply put, the speaker shows he or she has some evidence for the highlighted possibility, α, and the level of evidence depends on the sentence type. Rising

declaratives are used when the speaker has little to no evidence, rising tag interrogatives when the speaker has some or a lot of evidence, and falling tag interrogatives when the speaker has a lot of evidence.

Farkas and Roelofsen (2017) take this approach because it has some advantages in real-life discourse. Consider a teacher using a rising declarative to implement an echo question, or reversed polarity question (Koshik, 2002, 2005), to show to a student that the provided answer was not correct. In the approach proposed by Gunlogson (2008) the teacher would be making a contingent commitment to the expressed proposition. So not only is the student considered to have authority, but the teacher would also be committing to the wrong answer should the student confirm the wrong answer that was formulated in the echo question.[23] By taking a rising declarative as an inquisitive sentence type, this problem is averted. The teacher simply highlights the answer provided by the student, and adds $\langle \alpha, [\text{zero}] \rangle$ to his or her evidence list. The student can then infer that his or her answer was wrong, since the teacher would otherwise be asking a superfluous question.

To sum up: Farkas and Roelofsen (2017) propose that a sentence type can be informative or inquisitive, and that it can be marked or unmarked. All sentence types receive the same basic conventional discourse effect in which the speaker conveys the information expressed by the proposition, and steers the conversation towards a context in which both participants are committed to a state that supports the expressed proposition. Marked sentence type have an additional discourse effect that models that the speaker also has evidence for the highlighted proposition, or sentence radical, ranging from zero to high.

Discussion

Neither the proposal by Gunlogson (2001, 2008) nor the one by Farkas and Roelofsen (2017) deal with action formation and ascription as these are understood in CA. Nevertheless, both methods are aimed at modeling the semantics of linguistic structure in a way that supports the various pragmatic functions that these structures are used for. Given that structures like declaratives or rising boundary pitch do not have a one-to-one correspondence to certain actions, they attempt to understand what role these structures do play. I first show how these proposals could be applied, arguing that in their current state they fall short of providing a exhaustive explanation. In the next section, I then discuss

---

[23]This analysis is actually incorrect. The teacher's echo question is not an initiating declarative question and so Gunlogson's analysis does not apply. This does show that her approach is rather limited.

how these methods could be combined with CA to further our understanding of grammar in interaction.

An important starting point of both approaches is that they take syntax and prosody to make distinct contributions to the meaning of an utterance and that this meaning does not correspond one-to-one to certain actions. There are forms and there are actions, and while certain forms are frequently used to implement certain actions, that does not mean those forms do those actions. Any association between form and function is a result of certain underlying factors. So any form can be used for any action, but due to their semantic interpretation or conventional discourse effects, some forms may be more natural for some actions than for others.

Take for example the following excerpt:

(12)   [MidWest 2.4] (Heritage, 2012a, p. 8)

```
1   DOC:    Are you married?
2           (.)
3   PAT:    No.
4           (.)
5   DOC:    You're divorced (°cur[rently,°)
6   PAT:                    [Mmhm,
```

The physician's turn in line 1 is understood by the patient as a polar question, which she reveals by disconfirming in line 3. But why is it understood as a question? One account is of course that the physician uses a polar interrogative. But without the LFH that is merely an observation, not an explanation. What makes the physician use a polar interrogative to make his utterance accountable as a question, and as this particular question at this point in the interaction? And consequently what changes that in line 5 a declarative is the format chosen by the physician?

Gunlogson (2008) and Farkas and Roelofsen (2017) provide different explanations. For the polar interrogative Gunlogson (2008) would argue that by using a polar interrogative the physician conveys that he cannot be the source for this information and thus that the patient has to be the source, which means that he treats the patient as implicitly authoritative with regard to her marital status. Under the approach by Farkas and Roelofsen (2017) the physician puts the proposition expressed by the utterance on the table. Since it is a polar interrogative, this proposition has an inquisitive content of two alternatives which requires a response from the patient in which she commits to either being married or not.

The subsequent declarative does not fall under Gunlogson's (2008) framework as it is not an initiating declarative question. Farkas and Roelofsen (2017) on the other hand have no such limitation. Although the transcript is not entirely clear, it seems to be a rising declarative, which means that it functions as a question for the same reason the polar interrogative did. But it has the additional effect that the physician conveys that he has weak evidence for the patient being divorced. Indeed, the patient has only said she is not married, which means she could simply never have married or she could be a widow. So the rising declarative is the format that we would expect based on the proposal by Farkas and Roelofsen (2017) and indeed their proposal explains why the physician would choose a rising declarative here and not say a polar interrogative—that would suggest he is completely neutral—or a tag interrogative—the physician should not just assume that the patient is divorced simply because she is not married.

But an obvious shortcoming of both proposals, as well as of course of similar proposals in semantics, is that they do not ground their analysis in recordings of actual conversation. It is a point that is particularly noteworthy since Gunlogson (2001) and Farkas and Roelofsen (2017) suggest their approaches are empirical. While they use data *from the wild*, these data are not recorded and therefore cannot be subjected to re-examination nor to critical re-analysis. In effect the data are no longer the brief burst of talk produced in everyday life, but they consist of the written record of that burst of speech and the subjective judgments of the researchers about which alternatives would or would not have been felicitous in a given context. But such an approach to a phenomenon like language use is fundamentally problematic. As was pointed out by Sacks, by relying on their own judgments, the researchers inherently limit their potential understanding of the phenomenon they study:

> If a researcher uses hypotheticalized or hypotheticalized-typicalized versions of the world, then, however rich his imagination is, he is constrained by reference to what an audience, an audience of professionals, can accept as reasonable. That is to say, theorizing in that fashion has as one boundary on it that only those things can be offered which pass under some notion of believability. (Sacks, 1995, volume II, p. 419)

Actions are not a clearly defined set but, as Schegloff (2007) points out, they constitute a class of unknown size. And so we cannot limit our study to the language we ourselves, our friends, colleagues, and relatives produce and

judge to be adequate. To quote Sacks (1995, volume II, p. 420) again: "From close looking at the world you can find things that we couldn't, by imagination, assert were there." Consider for example such a phenomenon that one would only find by taking a close look at actual interaction: a topic proffer (Schegloff, 2007). Topic proffers are recurrently implemented with what would could be categorized as declarative questions. Take the following example:

(13)   Stolen, 2:18-25 (Schegloff, 2007, p. 176)

```
1   Mar:      But Ile:ne probably (0.8) is either at the
2             airport er waiting tuh hear fr'm eess
3             (0.7)
4   Ton:      O:kay.
5   Mar: ->   .hhhh So: yer ba:ck.
6   Ton: ->   Yah.
7             (1.0)
8   Mar: ->   I see. So you'll- you'll hear fr'm im,
```

After closing an arrangement-making sequence in lines 1–4, Marsha proffers a new topic in line 5. Her utterance has declarative word order, and is understood by Tony as making relevant confirmation. It is for all intents and purposes a question for these participants. Initially it seems unproblematic for the analysis proposed by Gunlogson (2008): the question is discourse-initial, but it should be clear for the participants involved that Marsha's commitment about Tony's whereabouts is dependent on Tony's confirmation, a contingency that could be said to be marked with turn-initial *so*. In other words, Tony is implicitly authoritative and Marsha asks a contingent question.

But that would skip over an important feature. They have just talked about their son moving back to Tony, as he was staying with Marsha while Tony was away. In other words, Marsha's commitment cannot possibly be dependent on Tony's subsequent ratification; she can be understood to already know that Tony is back, and so her utterance should in no way be considered questioning.

An argument could be made either (a) that Marsha's utterance is in some way not discourse-initial or is directly related to prior talk, or (b) that it is not doing questioning and so falls outside the scope of Gunlogson's proposal. But if Marsha's turn is not discourse-initial, what is considered discourse-initial becomes undefined and open to ad-hoc stipulations. This would clearly be undesirable in any semantic theory, because then anything goes. Similarly, from the data we have to understand Marsha's turn as doing questioning: Tony confirms and there is no evidence that contradicts such an analysis.

So claiming that Marsha is not doing questioning would either also be an ad hoc stipulation, or would require us to ignore the actual participants'

understanding of their own actions. But if our linguistic theory is not about what language does and means for the people who use it, then we have no actual theory of language. So at the very least Gunlogson's theory needs refinement.

The example does fit the proposal by Farkas and Roelofsen (2017): since Marsha can be taken to know that Tony is back, a falling declarative is precisely what we would expect. Alternative formats like a rising declarative or tag interrogative would suggest that she has evidence, but is not yet sure. So this topic proffer seems fine.

But consider a different type of action, again one that is not discussed by either Gunlogson (2001, 2008) or Farkas and Roelofsen (2017): second assessments that are implemented with a tag interrogative. In the following example two friends, Vera and Jenny, are talking on the phone. Vera's son, daughter-in-law, and grandchildren came to visit, but because Vera was not home, they stayed at Jenny's first. The excerpt is taken from a stretch of talk in which they are talking about one of the grandchildren. My focus is on line 35.

(14)   [Rahman:14:1–2] (G. Raymond & Heritage, 2006, p. 697)

```
33   Ver:   f→    =[Mindju 'eez good] Jenny, 'e wz mischeevious
34          f→    but w-'e wz good.
35   Jen:   g→    Oo 'e wz beautiful here [wuz↑n't'ee.=
36   Ver:                                 [↓Yes.
```

Vera in line 33–34 gives a positive assessment of her grandchild, to which Jenny responds with a similarly positive assessment. The design of her turn, however, is not what we would expect under the proposal by Farkas and Roelofsen (2017). Jenny talks about how the child behaved at her place; she does not just have evidence for that, in this conversation she is the only one who can possibly know how Vera's grandchild behaved at her place. While her utterance is not exactly neutral, we would have to argue that she is not committing to whether or not the child was beautiful, but that she leaves open both options and relies on Vera, who was not even there, to commit either way.

In other words, under the proposal by Farkas and Roelofsen (2017) Jenny's turn is not informative, it is inquisitive with a strong bias. But that is not in line with the actual context the participants are in. Vera is in no position to resolve the issue of whether or not her grandchild was beautiful, yet we would have to argue that she resolves the issue in line 36. As with the prior case, the only way to address this incongruence is by stipulating that in this particular case Jenny behaves as if she is not fully knowledgeable and relies on Vera to

confirm, as they are talking about Vera's grandchildren. But such stipulations mean anything goes, because we can always stipulate that in some discourse context for those participants it is not about actual evidence, but about how they position themselves. So this example shows that Farkas and Roelofson's analysis also requires refinement if it is to account for the meaning of linguistic structures.

A convergence of methods

From a CA standpoint both approaches are inherently problematic. They rely on a distinction between three levels of analysis: (1) the meaning of the words and the way they are put together, their *semantic content*; (2) the meaning of the sentence types, their *conventional discourse effects*; and (3) the intentions of the speaker, their *pragmatic discourse effects*. In other words, any sentence will have a fixed meaning that may get interpreted differently in different situations. But participants in social interaction are concerned with action first, not the meaning of an utterance (Heritage, 1984b, p. 139). Moreover, literal meaning as it would have to be assumed for these approaches is already context-bound, so no sentence will ever have a context-free meaning (Rommetveit, 1988). The fact that both Gunlogson (2001, 2008) and Farkas and Roelofsen (2017) include the boundary pitch in their sentence types, only exacerbates those issues (see T. Walker, 2014).[24]

More importantly, they assume that linguistic structure and meaning take primacy over action. The meaning that is ascribed to the linguistic structures is independent from the actions that those structures are used for. Farkas and Roelofsen (2017) even propose an analysis where structures have universal meanings, meaning both within the same language as well as cross-linguistically. But since language is used to implement action, so is its structure. This is why Schegloff (1996c, p. 110) argues that "one does not have "a grammar" for sentences (...) one has a range of grammatical resources," and he goes on to note that:

> Rather than starting with propositional forms and overlaying action operators, our primary characterizations need to capture the action(s) embodied in a burst of language. (...) There is every reason to suspect that grammar for talk implementing action is quite

---

[24]Note though that a formal semantic approach that desires cross-linguistic explanatory adequacy needs to include more than syntax in sentence type, as many of the world's languages do not distinguish between interrogatives and declaratives morphosyntactically, but may do so prosodically (Dryer, 2013).

different from grammar for talk expressing propositions. That we
may not yet have much of a clue as to what such grammar(s)
look(s) like does not change the suspicion, but may encourage a
sort of reaching that promotes the possibility of grammars rather
than *a* grammar. (Schegloff, 1996c, p. 113)

We should thus not assume one underlying grammar and build a semantic
analysis on it, and we should definitely not look for such a grammar to have
cross-linguistic implications. Instead we should consider that "language is first
and foremost a tool for [implementing actions in] interaction (Couper-Kuhlen
& Selting, 2001); language is not a discrete dimension of human cognition
but is used in service of the originally embodied actions that it has come to
augment and refine (see Tomasello, 2003, 2008). As conversation is "the natural
home of speech" (Sacks et al., 1974, fn. 1), we would expect that the design of
language is shaped both by and for the actions and functions that participants
concern themselves with in conversation (see also Couper-Kuhlen & Selting,
2001; Ogden, 2006).

That is not to say the approach used by Gunlogson (2001, 2008) and Farkas
and Roelofsen (2017) is misguided; the explanatory adequacy of their models,
even if they deal with limited and constructed evidence, shows that they could
very well have merit. But instead of interpreting them as universal semantic
theories, we should consider their value as particular grammars for particular
sets of actions.

In fact, Gunlogson (2008) could be said to be developing just that. Her
interest is in Initiating Declarative Questions; that is, questions that are imple-
mented in a specific position in the overall structural organization of interaction.
She is thus developing a grammar for specific actions—questions—in a spe-
cific sequential position. Indeed, her analysis relies on a crude version of the
adjacency pair (Gunlogson, 2008, p. 128):

(15)   A discourse move μ by an agent α is *contingent* upon a discourse con-
        dition δ if:
        a. δ does not obtain at the time of μ
        b. It is inferable in the discourse context that the update effected by μ
           is to be retained only if δ obtains the discourse move immediately
           succeeding μ

The effect of any contingent discourse move, be it a declarative question
or some other type of action, depends for its success on affiliation by another

participant. These discourse moves thus in a way project a next move by another participant, that is, such contingent discourse moves are FPPs. And while the criterion that such a next move has to be produced immediately is too strong—various actions can be inserted after an FPP—this is also recognized by Gunlogson (2008, fn. 18), who suggests that that demand can be relaxed. There is thus a clear convergence of methods here, and CA scholars would do well not to dismiss these findings out of hand, just because they are not grounded in strict CA principles.

Similarly, CA can learn a lot from the proposal by Farkas and Roelofsen (2017). The primary problem with the tag interrogative example is that it implements an assessment, not a request for information. It would seem that for assessments a tag interrogative is not used in the same way as it is for a request for information/confirmation. If we restrict the analysis by Farkas and Roelofsen (2017) to a smaller set of actions, if we start with action and not linguistic structure, it might gain us one of the grammars that Schegloff (1996c) suggests could be out there.

Formal linguistic methods such as these can provide us an understanding not so much of how participants in interaction make their actions accountable (Garfinkel, 1967), but why they do it in this particular way at this particular moment. It is part of the action-formation problem that CA has only sporadically concerned itself with, but it is obviously a question that is worth asking if we want a complete understanding of language.

1.4.3 Conversation Analysis

The final approach I discuss comes from the method applied in this dissertation: Conversation Analysis. Although CA has always been concerned with social action, and requests for information have received plenty of attention (e.g., Freed & Ehrlich, 2010; Koshik, 2005; G. Raymond, 2003; Stivers et al., 2010), how utterances come to be designed and understood as requests for information is an issue that has remained somewhat understudied until recently (but see Freed, 1994; Labov, 1970; Pomerantz, 1980; Schegloff, 1984). This is particularly surprising since CA has for decades shown that requests for information are part and parcel to institutional interaction. Whether it's news interviews (Clayman & Heritage, 2002a), press conferences (Clayman & Heritage, 2002b; Clayman, Elliott, Heritage, & Beckett, 2012; Heritage & Clayman, 2013), visits to a physician (Heritage, 2010; G. Raymond, 2010a; Robinson & Heritage, 2006), courtroom examinations (Atkinson & Drew, 1979; Sidnell, 2010), or any of a plethora of other institutional environments: they are almost all organized

largely through requests for information.[25] Understanding how this works would thus indeed seem to be foundational to an adequate description of these speech-exchange systems and an understanding of conversation in general (Heritage, 2012a). It is a problem that is obviously central to the field, as Schegloff (2007) also points out, but not one that is easily addressed.

A solution was recently suggested by Heritage (2012a; see also Heritage, 2013b, 2013a) who in a special issue of the journal *Research on Language and Social Interaction* argues, not unlike Beun (1989b) and Gunlogson (2008), that relative knowledgeability plays a large role. Although his proposal was initially embraced—Sidnell (2012, p. 59) in his response even wonders "how did we ever get along without this?"—it has recently become the focal point of some debate (Lynch & Macbeth, 2016a). In this section I will first present Heritage's proposal as it relates to the action-formation problem, and subsequently briefly address the current state of the debate. In closing I will discuss how his proposal is incorporated in this dissertation.

Epistemic Status and Epistemic Stance

The problem as it is presented by Heritage (2012a) is similar to Beun (1989b), Gunlogson (2001), and Farkas and Roelofsen (2017): If turn design cannot provide a definitive answer to the question of how participants distinguish between a turn that conveys and a turn that requests information, what then do participants use? His concern is thus not just with what makes a declarative turn understood as a request for information, but what makes any turn-at-talk understood as a request for information. Heritage seeks the answer in what he calls "the epistemic status" of the participants:

> When there is consensus about who has primary access to a targeted element of knowledge or information, that is, who has primary epistemic status, then this takes precedence over morphosyntax and intonation as resources for determining whether a turn at talk conveys or requests information. (Heritage, 2012a, p. 3)

This definition should not be understood simply as a precondition such as one might find in SAT. Heritage (2012a) is not concerned with stating conditions under which a request for information will or will not be felicitous. Instead

---

[25]Some speech exchange systems, such as emergency calls, may be said to revolve around a request for help. But even then the call-taker relies on a series of requests for information to determine whether help needs to be sent.

he argues that participants have epistemic domains (see Stivers, Mondada, & Steensig, 2011; see also Kamio, 1997; Labov, 1970; Labov & Fanshel, 1977; Pomerantz, 1980), personal areas of expertise, and these domains come with various rights and responsibilities (see Stivers et al., 2011). Most important among these for the action-formation problem is relative epistemic access. If the speaker is understood to have more access to some addressed piece of information—for example by being more knowledgeable or by having more authority—that speaker will generally be understood to be conveying information. And vice versa, if the recipient is understood to have primary epistemic status, the speaker will be understood to be requesting information.

This might seem to leave linguistics out in the cold. If differences between linguistic structure have no import on action formation, why then do participants design their turns in grammatically different ways? Heritage (2012a) argues that morphosyntax and prosody are primarily used to express what he calls "epistemic stance". Whereas epistemic status is stable and treated as more or less given, stance is a local expression of a speaker's epistemic position. It is used to fine-tune the "epistemic gradient" between the speaker who is in a relatively unknowing position ($K-$) and the recipient who is in a relatively knowing position ($K+$). Consider the following three sentences (Heritage, 2012a, p. 6):

(16)   Are you married?

(17)   You're married, aren't you?

(18)   You're married.

In each sentence the recipient has primary epistemic access and so when uttered each will be understood to be doing requesting information, but each indexes a different epistemic gradient. The speaker takes a more knowing stance in (18) than in (17), which in turn indexes a more knowing stance than (16). This can be visualized as in figure 1.1 (Heritage, 2012a, p. 7).

In general epistemic stance and epistemic status will be in alignment, meaning that a speaker lacking knowledge on some piece of information will take an unknowing stance, whereas speakers who already have a pretty good idea will take a relatively knowing stance. But stance can be used in a broader fashion. Speakers can use it to manage social relationships. G. Raymond (2010a) for example finds that health visitors in the UK will use a knowing epistemic stance to request confirmation and to treat a survey question as merely bureaucratic, whereas they can use an unknowing stance to request an actual answer and
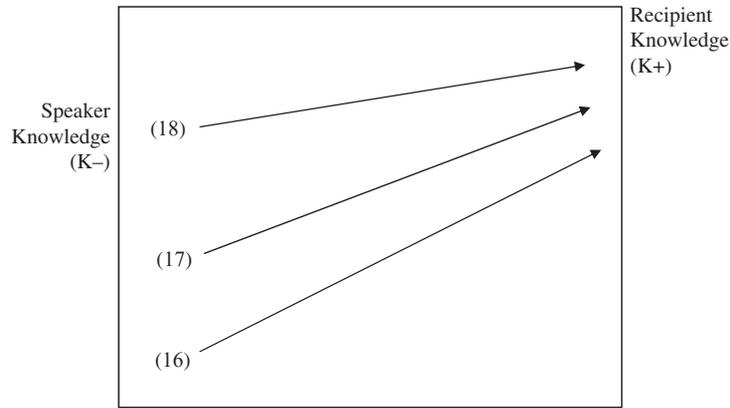
Figure 1.1: Epistemic stance of (16)–(18) represented in terms of epistemic gradient

thereby show a sincere interest in the mother. Similarly Heritage and Raymond (2005) find that in conversation, participants may take a more or less knowing stance, not because they are more or less knowing, but to deal with locally arising social exigencies. Additionally, congruence and incongruence between status and stance can be utilized to design actions; taking an unknowing stance when one is clearly in the know can be a means of doing challenging: for example, when a speaker who can clearly see what the recipient is doing asks *What are you doing?* (Drew, 2012).

The import of epistemic status in action formation is perhaps made most clear in those cases where a turn turns out to be truly ambiguous; that is, its uptake shows that it can in good faith be understood in one of two ways. Consider the following example where the speaker in lines 3–4 produces a declaratively formatted turn which is initially understood by the recipient to be the preface to a question but which is later understood to have possibly been a question itself:

(19)   [CNN Sate of the Nation 22nd March 2010: 8:56 EST]
       Conversation prior to the Congressional vote on health care reform
       (Heritage, 2012a, p. 11)

       1   Blitzer: .hh Kevin Madden you're- you're watching all of this and
       2   uh you are a Republican strategist.
       3   ->   Right now uh you realized of course the <u>D</u>emocrats are
       4   ->   going to <u>win</u>.
       5   (1.0)

```
6    Madden:   ->   Right.
7                   (.)
8    Blitzer:       So [then what.
9    Madden:   ->      [Yes. (0.2) Oh are you asking me or telling me.
10   Blitzer:       On this- on this issue of health care reform.....
```

The ambiguity arises here because while, as Heritage (2012a) points out, it was indeed clear that the Democrats were going to win the vote, Blitzer formulates his turn to address something that falls squarely in Madden's epistemic domain: what Madden may or may not realize. Possibly because the result of the vote was so obvious, Madden initially understands Blitzer to be simply making an assertion, building up to a question, but later realizes—indexed with *oh* (Heritage, 1984a)—that Blitzer's turn could also be doing asking.

And not only declaratives are susceptible to be heard as both conveying and requesting information. Consider the following example:

(20)   [KR:2] (Terasaki, 1976/2004, p. 202)

```
1    Mom:          Daddy 'n I have t- both go in different
2                  directions, en I wanna talk t'you about where I'm
3                  going (t'night).
4    Rus:          (Mkay,)
5    Gar:          Is it about u:s?
6    Mom:          (Uh) huh,
7    Rus:          <I know where yer goin,
8    Mom:          Whˆere.
9    Rus:          To thuh eh (eight grade )=
10   Mom:          =Yeah. Right.
11   Mom:     ->   Do you know who's going to that meeting?
12   Rus:          Who.
13   Mom:          I don't kno:w.
14                 (0.2)
15   Rus:          .hh Oh::. Prob'ly .h Missiz Mc Owen ('n Dad said)
16                 prob'ly Missiz Cadry and some of the teachers
```

Mom's turn in line 11 is initially treated by Russ as a pre-announcement, that is, preliminary to the doing of an actual announcement (see Terasaki, 1976/2004). By using the same question word as Mom, *who*, he provides her the opportunity to tell him who's going (Schegloff, 1988b). But as it turns out, Mom was not doing a pre-announcement, but requesting information (or possibly doing a pre-request), as she makes clear in line 13 by saying that she does not know. In other words, she is not a knowing participant who has

something to tell to Russ, but an unknowing participant, who wants Russ to provide her with information. Russ acknowledges this repair initiation with *oh*, and goes on to provide an answer.

What is thus crucial in this sequence is whether Mom can be seen to setting herself up to convey information or is requesting information. The design of her turn can apparently be used for both. As Terasaki (1976/2004) points out, Russ has just presented Mom with a riddle in line 7, and her turn in line 11 can therefore be understood as a return-riddle; this is indeed how Russ takes it up. But through his first riddle Russ has presented himself as knowledgeable on his mother's plans, and this seems to be what she is orienting to by requesting information (Heritage, 2012a). The sequential context thus provides different means of understanding line 11, and this ambiguity is only resolved when Russ has attributed the right status to Mom as either more or less knowledgeable than he is. While her turn design can definitely provide clues—had she said "Who's going to that meeting" her turn would no doubt be understood as a request for information, instead Russ now relies on its potential idiomatic function as a pre-announcement—to adequately grasp what action Mom is doing, to understand the relevance of her turn design, Russ cannot rely on turn design alone.

These examples show that the design of a turn is interpreted by its recipient in relation to the action that turn is understood to be doing. There is thus something of a two-way street between turn design and action ascription. Participants have to rely at least partly on the design of a turn to grasp what action a speaker may be doing: language is after all to a large extent the tool with which speakers implement social actions. A recipient cannot understand a speaker to be doing requesting information completely independent from how that request is implemented. But at the same time language alone cannot provide a complete answer, it can at best help guide the recipient given the specific sequential environment and social context. It is only once an action has been ascribed to a turn that the full import of its design can be adequately grasped. Turns are built incrementally and recipients are thereby consistently provided with opportunities to revise their projections of what action the speaker is implementing.

To sum up: Heritage (2012a) takes a fundamentally different view of the action-formation problem when it comes to turns that either convey or request information. Where prior research assumed that linguistic structure is the most important contributor, Heritage argues that participants rely on the socio-epistemic context: what they assume they and their interlocutors know and have a right to know. This, he argues, is in fact why many languages do

not have polar interrogative morphosyntax.[26] In this way linguistic structure is freed up for doing other things, such as "navigating the epistemic landscape" (Heritage & Raymond, 2012) or mobilizing response (Heritage, 2013a; Stivers & Rossano, 2010).

Epistemics as a hidden interactional order

The proposal by Heritage (2012a)—or proposals, as he also argued for the importance of epistemics for sequence organization (Heritage, 2012b)—is as Drew (2012) puts it a summary of a research agenda that was stimulated by Heritage and Raymond (2005) and G. Raymond and Heritage (2006). While there are of course discussions about knowledge in interaction to be found in lots of prior work, it is only since those two papers that epistemics has gained a central role in much conversation analytic inquiry. As Drew also points out, this agenda had become incredibly influential: "The stream of articles (...) is already proof of the importance and generativeness of what Heritage has been presenting and publishing on epistemics" (Drew, 2012, p. 61f.).

But despite its influence on contemporary CA research—or precisely because of it—the incorporation of epistemics in CA studies is not without its critics. In a special issue of the journal *Discourse Studies* (Lynch & Macbeth, 2016a), a series of articles was published that had as its aim to give what the authors in this special issue call the Epistemics Program the critical attention it was due (Lynch & Macbeth, 2016b, p. 494). As the analytic scope of epistemics is rather broad, ranging from action formation and ascription to sequence organization and recipient design, so too do the papers in this special issue deal with a whole range of analytic factors. As I am concerned primarily with action formation in this dissertation, I will limit my discussion to the response by Lindwall, Lymer, and Ivarsson (2016) and its uptake (Heritage, 2018; Lymer,

---

[26]While many languages indeed lack interrogative morphosyntax, the far majority—well over 80%—of the world's languages do have polar interrogative morphosyntax (Dryer, 2013). So Heritage makes a rather sweeping claim. Given the basic communicative functions posited by Tomasello (2008, p. 83ff.)—Requesting, Informing, and Sharing—linguistic structure may have arisen precisely to deal with action formation, but evolved to deal with the changing complexities of social relations, changes that were made possible by language. As Sidnell (2012) points out and Heritage (2013b) also acknowledges, there are plenty of languages where evidentiality is grammaticalized (de Haan, 2013), in other words where epistemic status and not just stance needs to be expressed (although Heritage (2013b, p. 393) argues, at least for Japanese (see Hayano, 2011), that this is still an expression of epistemic stance). The degree to which epistemics are "a deeply indigenous feature of human interaction" (Heritage, 2013b, p. 392) and its import for linguistic structure have so far not been adequately studied and are thus a prime field for future research.

Lindwall, & Ivarsson, 2017), as Lindwall et al. (2016, p. 501) focus on the question of "whether the empirical demonstrations really show that epistemic status is a fundamental and unavoidable component of the production and recognition of social actions."

Lindwall et al. (2016) take issue not with the problem that Heritage (2012a) investigates—the action-formation problem—but the way in which they perceive that Heritage approaches it. First, they align with the position that turn design alone cannot provide an adequate explanation for action formation, but they argue that Heritage has no consideration for the role of sequence. As mentioned in the previous section, Heritage (2012a, p. 3) argues that epistemic status takes priority over linguistic features of turn design, but Lindwall et al. (2016, fn. 4) point out that Schegloff (2007, p. xiv) gives a broader selection of possibly relevant criteria, such as the position in the sequence and the environment of the interaction (see Schegloff's definition in section 1.1). The result of this limitation they argue is that Heritage is primarily concerned with single turns, comparable to speech-act theorists, which is a step back because one of the benefits of CA's sequential analysis is that by freeing analysts from the prison of single utterances, we get a more complete picture of action formation.

The second issue Lindwall et al. (2016) raise deals with the recognizability of epistemic status. They argue contrary to Heritage's claims about the straightforward way in which participants deal with epistemic status, it is not clear how they prioritize one aspect of the environmental and epistemic context over another. And if it is unclear how participants come to their understanding, so too is it for the analysts: "The access, rights, entitlements and so on are established and presupposed state of affairs in virtually any conversation. However, there are few clues how we, as overhearing analysts, are to ground *our* claims to recognize these matters" (Lindwall et al., 2016, p 506).[27]

Third, they point out that epistemic status is severely limited in its use. Consider that Heritage (2012a) was concerned with how participants can distinguish between actions that request and convey information, and obviously those are not the only types of actions that participants do. Consider line 5 in the following example where Shelley asks what could be called a rhetorical question:

(21)   Debbie and Shelley (Heritage, 2012a, p. 23)

---

[27]The problem they raise is thus very similar to Clark's (1996) objection to the commonsense approach to Common Ground. But it is not just a problem for Heritage (2012a), Schegloff (2007) similarly includes the environment as a possibly relevant feature for action formation, and it is similarly unclear how he would prioritize one aspect of the environment over another.

```
1   Shelley:    So: I mean it's not becuz he's- he's- I mean it's not
2               becuz he:'s not going it's becuz (0.5) his money's
3               not ¿ (0.5) funding me.
4   Debbie:     Okay¿
5   Shelley: -> So an' ↑when other time have I ever [done that?
6   Debbie:                                          [.hhh well I'm jus say:in'
7               it jus seems ou- you base a lot of things on-on guy:s.
8               (.) I do'know:, it just- a couple times I don- I don-
9               .hh it's not a big deal.
```

According to Heritage (2012a), because Shelley has epistemic primacy—her utterance deals with her own prior behavior—her turn is not understood by Debbie as information seeking, but as challenging, complaining, protesting, or something or this kind. Debbie also treats it as one of the latter by backing down, first with *I'm jus say:in'*, subsequently mitigating *It jus seems*, and finally her claim that *It just a couple times* and that *It's not a big deal*. But, as Lindwall et al. (2016, p. 512) point out, epistemic status does not explain why Debbie understands Shelley's turn as a challenge-type action, only why she does not understand it as requesting information. They surmise that Debbie has to rely on its turn design and sequential environment, but if that's the case, she no longer would need to figure out whether or not Shelley is requesting information as a separate issue. Furthermore, they argue based on earlier parts of the conversation where Shelley produces a similar turn and actually gets an answer, that it's not necessary true that she has epistemic primacy.

Based on Heritage's (2018) response to the special issue, it seems that much of the critique by Lindwall et al. (2016) is based on a misunderstanding of his proposal—although in a subsequent response Lymer et al. (2017) argue that there are clear differences between the proposal made in Heritage (2012a) and (2018). Heritage (2018) argues that the data show that sequence is given its rightful place. In the cases presented in Heritage (2012a, 2013b, 2013c) epistemic status can frequently be grasped from preceding talk, and this is in fact what the participants orient to. It it thus not a valid objection to say that "epistemic status is established and managed sequentially" (Heritage, 2018, p. 28). Recognizability is not primarily a concern, but is indeed frequently out there in the preceding sequence.

Their disagreement on the use of epistemic status also rests on a misunderstanding of the goal of Heritage's proposal. It is not the case that in examples like (21) Debbie deals with epistemic status first before moving to a correct understanding of Shelley's action. Consideration for status is only part of the action-ascription process: "The recognition of whether one is being asked or

told something is *part of the process by which one produces and understands a wide variety of action types* (...) prior to framing a response" (Heritage, 2018, p. 32). While epistemics are thus omnirelevant—which as (Heritage, 2018) argues should come as no surprise since many actions such as requests and offers are implemented with questions as vehicles (see Schegloff, 2007)—they are not omni-determinate. Participants have due consideration for all aspects that may be pertinent to action formation.

The point of discussion is thus whether epistemic status is necessary for the process of action formation. Lindwall et al. (2016) argue that if status can be grasped from the sequence, it is not a notion that adds anything to our analytic toolbox. Indeed, they fear it places a heavy burden on what they perceive to be ad hoc stipulations, resulting in less rigorous analyses. But as Heritage argues, participants rely specifically on epistemic status and so it is an important additional tool: "Part of the recognition of an action will arise from deciding how what is being talked about is positioned relative to the epistemic domains of the speakers" (Heritage, 2018, p. 32).

Discussion

As I have already discussed some of the responses to the role of epistemics for action formation in the previous section, I will provide only a short discussion here, focusing on how it relates to the proposals discussed in sections 1.4.1 and 1.4.2, before explaining how epistemics is used and understood in this dissertation.

Based on the brief summary presented of Heritage's (2012a, 2013b, 2013c) proposal we can see obvious similarities with those by Beun (1989b) and Gunlogson (2008).[28] Beun (1989b) posits that for an utterance to be a declarative question, the hearer needs to be reciprocally understood to be the Expert, and similarly Gunlogson (2008) argues that for any utterance, not just one with declarative word order, to be a question the recipient needs to be recognized as

---

[28]Farkas and Roelofsen (2017, p. 278) take a completely different perspective: they argue that "epistemic authority" cannot explain various of their cases. As they do not cite particular work, it is unclear whether they are critiquing epistemics as it is used in CA, or Gunlogson's (2008) implicit authoritativeness. While they do talk about "assumed knowledgeability", which suggest they take issue with CA's take on epistemics, they seem concerned only with epistemic stance as it is encoded in the design of turns, not epistemic status. Furthermore they rule out epistemic authority as a possible account for action formation because it would make some cases felicitous when they should not be. But CA, like participants in the conversation themselves, is not concerned with whether or not an action is felicitous. Since it is unclear whether they misunderstand CA or deal with an entirely different theory altogether, I leave their critique aside.

implicitly authoritative.

Clearly three very different methods—Speech Act Theory, Formal Semantics, and Conversation Analysis—converge on very similar proposals. Heritage's (2012a) proposal, however, offers advantages for an interactional analysis that those by Beun (1989b) and Gunlogson (2008) do not. First, unlike Beun (1989b) who deals with the recognizability of declarative questions only, Heritage proposes a broader theory of requests for information, indeed of the role of knowledge for action formation in general. Where Beun's analysis does prove valuable is, somewhat ironically, in his proposal for a complex function structure. As I argued, we would need a sheer infinite set of possible function structures and orderings of preferred interpretations. But those values can be understood as guiding the participants' grasp of who has primary status, indeed as expressions of that relation. It can help us understand how expertise is indexed in linguistic structures. This is also what Beun aims for: the function structures are supposed to help make clear who is Expert on some issue. All we would need to do is drop the assumption that such functions are in a sense universal and deterministic.

Although Gunlogson (2008) approaches the action formation problem from formal semantics, her analysis takes a dynamic approach to discourse and is thus at least somewhat interactional. She posits contingent actions and relies on implicit authority which needs to be mutually recognized. In a sense she attempts to formalize the adjacency pair as well as epistemic status in a more formal grammar. She does not quite succeed though: contingent actions are a more crude version of the adjacency pair, and implicit authority is somewhat cognitive; although the status as an implicit source must be inferable from the context, context includes an indeterminate set of unspoken and inaccessible assumptions. In fact, it should not be clear from the context who is the source, as that would mean one of the participants is already committed. So sequence is given a place, but not as prominent as in CA.

That is not to say that epistemic status is necessarily the perfect solution to the action formation problem—Heritage (2018, p. 34) says as much: "whether this claim [that epistemic status could always override (...) syntax and intonation as a basis for producing and recognizing assertions and questions] is true or not will probably take quite some time to sort out." I offer two brief points to consider, before formulating how I understand the role of epistemics in interaction in this dissertation.

The first deals with the recognizability of epistemic status. Heritage argues that status is typically a settled matter for the participants, and particularly that "the thoughts, experiences, hopes, and expectations of individuals are treated

as theirs to know and describe" (Heritage, 2012a, p. 6; see also Heritage, 2011). Consider, however, the following extract from a conversation between two sisters. Lisa is doing high school exams and will be going on a trip to Barcelona with her mother after she graduates.

(22)   BE1 02:34.1–02:41.3

```
01  Lis    °( [ )°
02  Fle       [°(ga ik)°
                go I
              [°(will I)°
03         (1.1)
04  Fle -> wat  ga ik dan  doe:n?=
           what go I  then do
           then what will I do:?=
05      -> =ga ik dan  iets     met  papa doen.
            go I  then something with dad  do
           =will I then do something with dad.
06         (1.0)
07  Lis    nee ↑jij hebt dan  nog  gewoon college,
           no  you have then still just   class
           ↑no you will still have class then,
08         (0.3)
09  Fle    °oh°.
           °oh°.
```

In lines 4–5 Fleur asks what she will be doing when Lisa is in Barcelona with their mother, whether she is going to do something with dad. Although one might argue that Fleur has primary rights to know and talk about her own future plans, Lisa treats Fleur's turn as a request for information by answering and explaining why she will not be doing something with dad, and Fleur's subsequent *oh* treats that informative response as adequately dealing with her turn in lines 4–5 (Heritage, 1984a; Schegloff, 2007). So clearly both treat Lisa as having primary rights here to talk about Fleur's future plans, or at least her plans as they relate to Lisa's planned visit to Barcelona.

One can look at this example and argue that for the participants it is at least clear that in this case Lisa has primary rights. That there is nothing salient in the data that confirms this for the analyst is not something Heritage (2018) is concerned with. It is after all a category for the participants and they have no problem with it. Indeed, if we take a look at the talk that precedes this interaction, we can ground the analysis in the participants' behavior. Lisa has not said when she will be going to Barcelona, in fact, Fleur initially displays no recognition of the trip whatsoever. So we can point to evidence in the sequence

that confirms the participants' status—whether that is or is not a useful way of looking at it is, as Lymer et al. (2017) argue, another matter.

This excerpt is not meant to disprove—or prove—that epistemic status plays a part in action formation, but as a cautionary tale. In the special issue on epistemics, Heritage (2012a, 2012c) and some of the responses (Clift, 2012; Drew, 2012) warn that epistemics should not be a go-to resource as that would rob it of its "empirical bite". This example doubly confirms that. If we were to sail blind on the idea that people's plans and experiences are theirs to know and talk about, then we would have a hard time explaining why Fleur's utterance gets treated as a request for information.[29] If instead we keep the importance of sequence in mind and see epistemics not as something more basic than turn-taking or sequence organization (cf. Sidnell, 2012), but as merely part of the process of action formation and ascription, then its value need not be lost.

That is in fact how I take epistemics and its relation to action formation and turn design to have been intended. Any turn-at-talk is designed to deal with the local exigencies of the interaction (Mazeland, 2013). Epistemic status does not transform an utterance that would otherwise be a request for information into an assertion or some other type of action, but it contributes to making an utterance understood as a particular type of action. In cases where relevant, such as requests for information, the design of a turn reflects, among a whole bunch of other stuff, the speaker's view of the relative access he or she and the recipient have to the state of affairs under discussion. By projecting a specific type of response, the speaker conveys his or her view of the relationship the participants have.

So in case of doing a topic proffer (see chapters 2 and 3) a declarative is immediately and unproblematically understood as a topic proffer; conveying information is not part of the picture. The choice for a declarative over an interrogative reflects the speaker's view of the relationship in which he or she knows and is allowed to know something about the recipient. Both formats are used to invite a telling, but with a declarative a speaker claims to know about the preferred news while with an interrogative a speaker leaves the options for both good and bad news open.

It is thus something of a misnomer to say that epistemic status overrides grammar, as grammatical constructions such as interrogative morphosyntax and rising boundary pitch have no invariant inquisitive meaning. Heritage (2013c,

---

[29] A simple account would be to say that here morphosyntax trumps epistemic status; Heritage (2012a, 2013b, 2013c, 2018) recurrently allows for that possibility. But that would only increase the risk of ad hoc stipulations: as analysts we can always say one or the other is given priority depending on what suits our analysis best, which would completely rob epistemics of any bite.

p. 569) also takes note of this when he says that such concrete meaning would be misplaced given its variability in different contexts. Grammar is just a tool for action formation. What makes forms like declaratives with a falling boundary pitch special is not that they request information or confirmation when they should not, but that they are used in specific social environments for specific interactional functions, as the various chapters in this dissertation attest.

1.5   Contents of this dissertation

At the start of this chapter I presented what could be considered a minimal pair:

(1)   The door is shut.

(2)   Is the door shut?

Both (1) and (2) can be used to do questioning, and so we cannot account for these functions based on the syntactic format of either. Neither, as was pointed out, can we rely on that other omnipresent action-ascriber: prosody. An utterance can do questioning with interrogative and declarative syntax, and with rising and falling boundary pitch. How then are we to account for the action status of either utterance if they are to be used as questions?

As will be clear by now, we cannot provide such an account without the context of these utterances. Sentences do not do actions, participants do actions, and they do so in an extremely rich context on which they continuously rely. Speakers design their actions to fit the exigencies of the interaction (Mazeland, 2013).

Furthermore, and perhaps more importantly, it is not clear what it means for a speaker to be asking a question. As pointed out by Schegloff (1984, p. 30), *question* is a commonsense category, not a technical one. While we have an intuitive idea what a question is—an action by which an unknowing speaker requests information of a knowing recipient—this cannot possibly cover all the types of actions we would be interested in if we were to study declarative questions. It is deceptively vague, and therefore far too broad: Many actions that look like they are doing questioning, are in fact concerned with some other action, such as inviting, requesting, or proffering a topic (Schegloff, 1984, 2007; Sidnell, 2017a). At the same time it is still too narrow: Declaratives are often used when a speaker could be expected to already know the answer (Heritage,

2012a; G. Raymond, 2010a), and so cannot be requesting information.[30] The very notion of declarative question under this definition would be something of an oxymoron.

So the question as it was put at the start of this chapter—If polar questions are made recognizable with a polar interrogative, how do recipients understand an utterance with declarative word order as a polar question?—is not the right question to ask. The approach in this dissertation is therefore to treat what are commonly considered declarative questions as actions that have a family resemblance (see Wittgenstein, 1958, p. 32). This means that instead of looking for declarative questions as if they were some clear-cut category of action, this dissertation is concerned with utterances in which participants produce first pair parts with declarative word order in which they make relevant a next action in which the recipient's primary concern is to affirm or confirm the state of affairs as formulated by the speaker.[31] This formulation attempts to capture that these actions function as what are traditionally considered polar questions, and to exclude such actions as invitations or requests. Note that this is not, and cannot be an exhaustive definition. It is used merely to guide the selection of the specific phenomena. It would likely be entirely unsuited for any formal coding process in service of statistical analyses.[32]

As we cannot assume the existence of declarative questions, the goal of this dissertation is obviously not to provide an analysis of how declaratives come to implement questions and how those differ from interrogatives. Instead, by considering a small subset of yes/no-type initiating actions with declarative syntax, or YNDs (G. Raymond, 2003, 2010a), and comparing them to similar actions with polar interrogative word order, or YNIs (G. Raymond, 2003, 2010a), I attempt to shed some light on the procedural aspect of action formation and ascription. The major points are that actions are understood in a specific sequential environment and that the action status of an utterance is an interactional accomplishment. In fact, action is always open to (re-)negotiation.

In chapters 2 and 3 I focus on the importance of sequential positioning for the process of action formation and ascription, that is, the sequential understanding of action. First I show that YNDs can be produced, and understood to be

---

[30]A more fundamental point of critique is that recipients cannot know what speakers knows; they can only assume and infer based on the presented evidence.

[31]State of affairs is not synonymous with propositional content. The latter presupposes a literal, context-independent, and grammar-independent meaning, and as should be clear from this chapter this is not how participants come to an understanding of turns at talk.
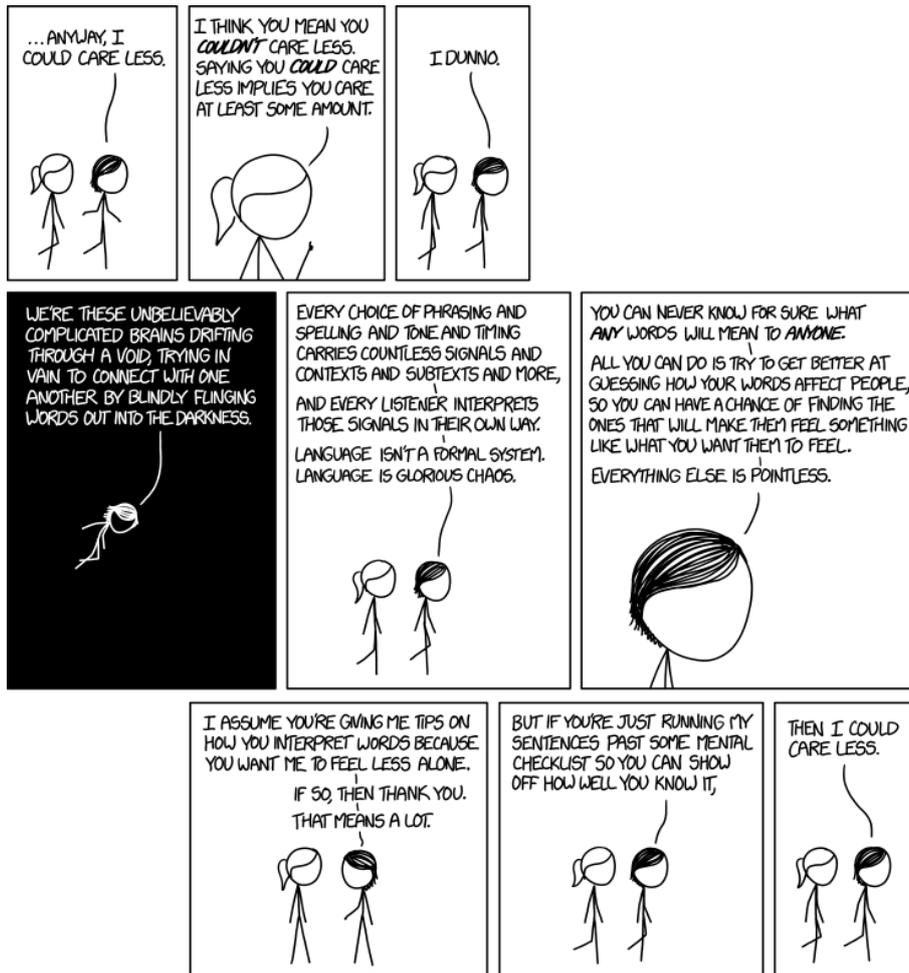
[32]No attempt at establishing its suitability by determining the Kappa coefficient of a sample— that is, intercoder reliability—has been made.

produced, to not just request confirmation or affirmation, but also some form of elaboration. This understanding is not, or at least not exclusively, based on the linguistic design of these YNDs. Indeed YNDs that request confirmation and YNDs that request elaboration do not differ morphosyntactically or prosodically in a clear way, but they differ in their environment of use. In environments of topic attrition where participants have to either start a new activity, re-open a previously closed activity, or move towards conversational closure, YNDs in which the speaker formulates a previously unaddressed matter will be understood to make relevant more than confirmation. That this is attributable to their sequential position and not some aspect of their design is supported by comparing them to YNIs in similarly topic-shift implicative environments. Although these YNIs will be understood to implement a different type of topic proffer—differences in design do of course matter—they will nonetheless also be understood as topic proffers.

In chapters 4 and 5 I focus on the role of grammar in interaction by comparing two specific morphosyntactic practices: *oh*-prefaced YNDs and *oh*-prefaced YNIs. Both practices are understood as addressing a potential breakdown in intersubjectivity, and by being *oh*-prefaced speakers use both to claim that their realization that there was a problem took place then and there and was touched-off by the immediate prior turn (Heritage, 1984a; Jefferson, 1978). But where YNDs are used by speakers to claim that they now understand after having either not understood or misunderstood at some prior point in the conversation, YNIs are used to convey that the prior talk by the interlocutor was not in line with the speaker's prior beliefs. Both therefore project different courses of action: an *oh*-prefaced YND makes relevant only confirmation, but *oh*-prefaced YNIs also invite some form of reconciliatory information.

Finally in chapter 6 I address the procedural nature of action, arguing that action is an interactional accomplishment and that in any adjacent turn recipients can not only display an understanding of the prior turn, but actively ascribe an action to it, possibly even recasting the status of their own prior talk. That is not to say that utterances need to be taken up for them to have implemented an action, but for an utterance to have a function for the participants in the interaction, both participants need to somehow converge on their understanding of it. In general this is entirely unproblematic, and the status of an utterance as a particular type of action is treated as salient. In these cases simply foregoing an opportunity to initiate repair after next turn (see Schegloff, 1992) confirms that it was adequately understood (Robinson, 2014). But action is not given, it is continuously open to negotiation, and therefore has to be accomplished collaboratively. It is through every next turn at talk that participants build and

maintain their architecture of intersubjectivity (Heritage, 1984b; Rommetveit, 1976).

I literally could care less.

Image courtesy of xkcd.com