THE UNIVERSITY
of EDINBURGH

IT'S ALL ABOUT THE INTERACTION
—
LISTENER RESPONSES AS A
DISCOURSE-ORGANISATIONAL
VARIABLE

Mirjam Elisabeth Eiswirth

The University of Edinburgh

A thesis submitted for the degree of Doctor of
Philosophy

August 2019

Part I

Front matter

Abstract

Language is humanity's key tool for communication. This entails the fact that it generally occurs in interaction between two or more individuals. However, we do not yet have a theory of language variation and change that integrates our understanding of the interactional nature of language with the variationist analysis. This is especially challenging for variables above the level of the phoneme, but also applies to phonological variables that are impacted by the interactional context they stand in.

This thesis focuses on Listener Responses, a variable above the level of the phoneme, and presents a theory and methodology of sociolinguistic variation that allows us to develop (1) interactionally sensitive definitions of discourse-level variables, the envelope of variation, and to quantify them in an accountable manner, (2) coding schemes which situate the function-based variants in the interactional structure and thus allow for an analysis of structural constraints on variation, and (3) a way of applying inferential statistics to variation based on structural as well as social variables. With respect to phonological variables, this thesis shows how the level of (inter)action relates to the actual realisations we observe.

This is done based on the example of Listener Responses as a discourse-organisational variable, and gender as a social variable. Listener Responses are defined as all the things Listener can do without taking over the floor. Their frequency is thus quantified relative to the number of words in the longer stretch of talk produced by the main Speaker. In the data at hand here, cross-gender accommodation is observed, with female Listeners decreasing their response frequency when listening to men, and male Listeners increasing theirs when listening to women. Next, a taxonomy of Listener Response actions is developed based on existing interactional literature and a close structural and interactional analysis of the data. Seven Action Types are proposed, and used as a coding scheme in the next two analysis chapters. The third analysis chapter shows variation in the frequency of the individual Action Types based on Speaker and Listener gender. There is an important structural constraint on variation located at

the level of interactional structure: those Action Types that are strongly predicated by what the main Speaker does in the segment preceding the response are more strongly influenced by the main Speaker, while the Listener has a greater impact on those that are not constrained by the preceding segment. Both the first and this analysis chapter draw on zero-inflated poisson regression analysis as a useful tool for the analysis of variation in frequency. The final analysis chapter looks at the relationship between the different action types and the actual linguistic realisation of the utterance, thus linking the discourse-level to the phonetic and prosodic level. It demonstrates that the linguistic realisation of any Listener Response is tailored to the talk that has preceded it on all levels of linguistic structure, and that prosodic and lexical shape need to be considered together, particularly for lexical items that can be used to do different actions.

Overall, this thesis contributes to sociolinguistic theory and methodology by presenting a way of integrating interactional and variationist analyses from the definition of the variable, the envelope of variation, an overall frequency operationalisation, describing and defining the variants, to exploring the link between interactional function and linguistic realisation. It can be extended to other variables, both linguistic and social.

Lay Summary

Scholars looking at how language varies and changes usually focus only on what a whole group of people does, or at how one person uses language in different ways. However, as individual speakers we use language to talk to and interact with other people. This does not usually play a role in the analysis of how language varies, but it is very important, especially for small signals like 'mhm' or 'okay' that we give when we listen to someone. I call those contributions 'Listener Responses'. Depending on whether the other person is looking for a word or telling us something surprising, different types of Listener Responses are appropriate – here, filling in the word, or signalling surprise. On the other hand, what we say and do as listeners influences how the other person continues talking. If we have given them the word, they can continue their turn. If we signal surprise, they often tell us a bit more about the thing we have marked as surprising.

In this thesis, I bring together the analysis of what participants actually do in their conversations and how they organise their speaking and listening through Listener Responses with quantitative analyses of language variation. I describe what Listener Responses are in the context of the interaction, and show that when we want to know *how often* any Listener (or group of Listeners) produces responses, we need to look at them relative to how much talk they have been listening (and thus responding) to (chapter 4). I then show that Listeners can do lots of different actions and that these are best described by looking at what the main Speaker does before and after the response (chapter 5), and how we can quantify how people (or groups) vary in how often they do these different responses. Again, we need to look at what the Speaker does before and after the response to make sense of this variation (chapter 6). Finally, I also look at *how* Listeners do these responses, which words they use and how they change their speech melody (chapter 7). Here we can see that what Listeners do and how they say it always matches what the Speaker has just said, either in terms of the words used, or in terms of the speech melody. We can also see that sometimes *how* a word is said is more important than the word itself to signal to the other person what we mean (just think of all the

different ways of calling out someone's name, and how that changes depending on why you are calling them).

Overall, we can see that what the Speaker does is really important to how many and which responses the Listener gives, and what these responses look like. There are some small differences between male and female Speakers and Listeners, but because there are more female than male participants in this study we cannot say that this applies to other men and women, too.

This thesis is important for academics studying language variation and change, and scholars interested in how we actually interact with each other, because it shows how we can bring both approaches together and how important it is to take into account that as people we usually use language in interaction.

Declaration of own work

I declare that this thesis has been composed solely by myself and that it has not been submitted, in whole or in part, in any previous application for a degree. Except where stated otherwise by reference or acknowledgement, the work presented is entirely my own.

Dedication

To my family
without whose unconditional love and support
I would not be where I am today.

Acknowledgements

This book represents a four-year journey into uncharted territory. It involved not only mapping the space, but also developing a whole network of roads and other infrastructure. Such a journey is simply impossible alone, and I would like to thank the people who have helped me populate this map, build the infrastructure, and make it come alive.

I am grateful for the advice and support of my supervisors, Lauren Hall-Lew, Joseph Gafaranga, and Joe Fruehwald. Thank you for having such an open mind when it comes to integrating two approaches and methods that often seem irreconcilable, for our discussions, your reading suggestions, and comments on my many paper and chapter drafts. I'd also like to thank Richard Ogden for hosting me as a visiting doctoral researcher at the University of York.

Thank you to the people living with Type 1 Diabetes who have participated in this study, allowing me to analyse their conversations and stories. Without your voices, this thesis would not exist. And thank you to Daisy Smith, who helped me collect this data, learn statistical methods and R, and develop my first big non-linguistic quantitative data analysis project.

'It's all about the interaction' – not only this thesis, but also how it came about. A big thank you to all the people I have interacted with in the process of developing my map and infrastructure. Thank you to colleagues at the Glasgow University Laboratory of Phonetics for the opportunity to discuss my work with you at different stages, and to Jennifer Smith for granting me access to the Buckie corpus for my pilot study. Thank you to audiences at DiPVaC3 in Helsinki and the English Linguistic Circle in Trier for helpful discussions of the coding scheme and quantification process, and at ICLaVE10 for ideas on next steps. Thank you to friends, colleagues, and students, who have given me feedback on different aspects of this thesis at different stages: the Sociolinguistics Reading Group (especially Zac Boyd, Claire Cowie, and of course Lauren Hall-Lew and Joe Fruehwald) and our EMCA group SEDIT here in Edinburgh (especially Sue Widdicombe – who also acted as an annual review board member for me twice –, Eric Laurier, Steve Kirkwood, and Eve Mullins), as well as the York CA group, Julia Moreno for great discussions

Contents

List of Tables

List of Figures

Part II

Setting the scene and building the foundations

Chapter 1

Introduction

Caminante, no hay camino,
se hace camino al andar.

–

Wanderer, there is no path,
your walking will make the path.

Antonio Machado (1912)

Language and interaction are inextricably intertwined: Language is the key tool for communication, and the majority of language occurs in direct interactions between two or more people. Whatever we say only means something and has an effect, because it is taken up by somebody else. Despite this interactional nature of language, we do not yet have a sociolinguistic theory or methodology that allows us to approach language variation while integrating the interactional situatedness of language production.

Instead, there are somewhat disparate traditions of scholarly work: The first looks at language variation with a focus on (quantitative) patterns of variation in the linguistic production across individuals without paying much attention to the interactional structure. The second focuses on the structure of interaction without paying much attention to distributional patterns. This PhD thesis proposes a theory of language variation that integrates quantitative variationist approaches with qualitative interactional analyses, based on an exemplary analysis of the discourse-organisational variable Listener Responses and the social variable gender. This methodology can then be transferred to other social and linguistic variables.

The quantitative study of sociolinguistic variation started with Labov's socio-phonetic studies of Martha's Vineyard (Labov 1963) and New York (Labov 1966b), and has since been extended to include variables above the level of the phoneme. Of specific interest here are discourse-pragmatic variables (Lavandera 1978; Dines 1980; Pichler 2013b). However, a number of challenges to modelling discourse-

level variables following the principles outlined by Labov (1966a) in his pioneering paper remain. They are as follows: (1) How can we define and accountably quantify the (frequency of) occurrence of a discourse-level variable? (2) How can we develop an interactionally rooted coding scheme that takes into account interactional structure as an internal conditioning factor? (3) How can we accountably quantify variation with respect to these functional variants, ideally with inferential statistical models? (4) How does this relate back to the level of linguistic realisation, particularly phonetics and prosody, i.e. how can we integrate an awareness of interactional structure into our analysis of phonetic variation?

Based on the exemplary case of LISTENER RESPONSES and gender, this thesis shows that it is possible to apply the analytic rigour of the Labovian definition of the phonological variable to discourse-level variables. LISTENER RESPONSES are chosen as a discourse-organisational variable for three specific reasons: First of all, there is a fairly large body of work on the phenomenon, from both qualitative and quantitative perspectives. Second, this work exemplifies the challenges that have been described for work on other variables above the level of the phoneme. Third, the final step of the analysis allows us to link discourse-organisational variation back to the specific linguistic – phonetic and prosodic – shape of the utterance, and thus to the traditional sociophonetic variable. This opens up avenues for integrating this interactional approach into sociophonetic analyses.

With respect to the treatment of GENDER as a social variable in this thesis, I would like to quote Goffman (1981: 1): 'So I ask that these papers be taken for what they merely are: exercises, trials, tryouts, *a means of displaying possibilities*, not establishing fact' (my italics). In other words, GENDER is chosen as the social variable to support these theoretical and methodological points, using a strategically essentialist definition (Meyerhoff and Ehrlich 2019). It serves as a stand-in for any other social variable like age or ethnicity (both macro-social categories with similar challenges) or more locally relevant identities. GENDER was chosen over any of those other social variables even though my sample is not perfectly balanced, because there is a large body of work I can draw on, both with respect to language and gender in general and LISTENER RESPONSES in particular.

I argue and demonstrate that LISTENER RESPONSES are an intrinsically interactional phenomenon and need to be defined, quantified, and analysed with respect to the interactional structure in which they occur. This means we need to consider the linguistic material or actions *surrounding* the LISTENER RESPONSE as essential conditioning factors on the realisation observed. Hence, the behaviour and speech of the *addressee*, here the turn-holder, which I will refer to as (main) *Speaker* with a capital S, are crucial to our analysis of the behaviour and speech produced by the *Listener* (also with a capital letter), i.e. the person uttering the LISTENER RESPONSE.

In the following, the SMALL CAPS will be referred to in SMALL CAPS and the *variant* in *italics*. Because this shifts as I step through the different levels of the analysis I will point out the VARIABLE and the *variants* at the start of each chapter.

The four analysis chapters of this thesis tackle the four key questions in turn: The first big question is how to do justice to the 'principle of accountability' laid out by Labov, which requires us to count *all* possible realisations of the variable, including 'null realisations'. Many discourse-level variables are characterised at least in part by their optionality, which makes it impossible to count absences. This begs the question how to operationalise their frequency of occurrence, and how to statistically model them. In response to this challenge, I present an interaction-based definition of LISTENER RESPONSES: they are all those contributions by the Listener which support the ongoing talk of the Speaker without laying any claim to the floor. This means they are structurally couched in the main Speaker's talk. We can only count those responses that do occur, not the absences. Nevertheless, it is possible to quantify their frequency in an interactionally accountable manner: relative to the number of words in the ongoing stretch of Speaker-talk they are responding to. Statistically, this is done using Zero-inflated Poisson-regression models. This 'ongoing stretch of Speaker-talk' is what I operationally define as a turn: the number of words from one Speaker-change to the next, based on what I had annotated as 'Listener Response' or 'talk'. Thus, a 'turn' is all the speech produced by one party until the other participant produces talk that is *not* a LISTENER RESPONSE any more.

In chapter 4 we find that female Listeners respond at the highest frequency when Listening to other women, while men Listening to men produce the fewest responses. There seems to be cross-gender accommodation in the mixed-gender dyads, with women decreasing and men increasing their LISTENER RESPONSE frequency. However, in the inferential statistical model the only strongly significant effect is turn length, followed by Speaker-gender as a marginally significant factor. This is interesting in so far as what the person being responded to does, and who they are appears to be more important in predicting the number of Listener Responses than characteristics and actions of the person actually producing those responses.

The second challenge is developing an interactionally sensitive, emically valid description of the functional variants, i.e. all the different actions Listeners can do. So far, these descriptions and definitions of variants are situated on varying levels of structure and mostly superimposed by the analyst rather than drawn from the interaction itself. We need (1) a clear definition of the level of structure at which we define these variants, and (2) an emically valid, interactionally sensitive definition and description of any given set of variants. It is possible to do so for LISTENER

RESPONSES by drawing on previous Conversation Analytic work on Listening and critiques of the undifferentiated quantitative treatment of backchannels that 'not all [LISTENER RESPONSES] are created equal' (Drummond and Hopper 1993b: 175) (see also Schegloff (1982), Gardner (2001), and Sorjonen (2001)). The most important tool is the so called next-turn proof-procedure: participants in an interaction negotiate the meaning of any utterance turn by turn, and as analysts we can understand what any utterance has been taken (and agreed) to mean by looking at how the other person treats it in the next turn. This procedure will be explained in more depth later. It is one of the key methodological tools that can be applied to the analysis of other discourse-pragmatic variables, too, as we will see in the discussion. The seven *Action Types* presented in chapter 5 are based on the sequential structure of the interaction, responding to Schegloff's (1982: 88) call to include the preceding context into the definition of any given LISTENER RESPONSE. This qualitative interactional analysis and focus on the importance of the sequential structure draws our attention to two aspects that are essential to the third and fourth big question: LISTENER RESPONSES are shaped by the (preceding) context and create the (following) context. They are thus by definition situated in the interactional structure.

The third question relates to quantitatively analysing this variation. Without a definition of the variants rooted in interactional structure, it is not possible to model interactional constraints on their production. A further complication is that the variant produced is generally attributed only to the person producing it, without paying attention to possible interactional constraints. Chapter 6 takes up these seven *Action Types* as individual variants of LISTENER RESPONSES and shows that their frequency is constrained by interactional structure: drawing on the next-turn proof-procedure again, we can see that some Action Types from the taxonomy presented in 5 are directly made relevant by what the Speaker does in the preceding segment of talk, some are negotiated by both parties equally, and for some it is mainly the Listener who can decide to utter them or not. Consequently, it would be premature and simplistic to ascribe the responsibility for any given action exclusively to the person who ostensibly 'does' said action. Instead, as indicated previously, it is important to consider how this action has come about, and which party had how much impact on it. This observation alerts us to the necessity of taking the preceding segment of talk into account as a structural constraint when analysing the relative frequency of each Action Type. In the quantitative analysis overall, Listener gender has a greater impact on which *Action Types* get done than Speaker gender. However, when we take into account that the preceding segment of talk constrains which actions can get done next more strongly for some responses than others, we find that Speaker and Listener gender interact with

these categories. Listener gender is a more important predictor for the relative frequency of those actions that are not directly made relevant in the preceding Speaker-talk, while Speaker-gender has a noticeable effect on the frequency of those actions that are directly made relevant by the preceding segment of talk produced by the Speaker. Here, just as for overall frequency, Zero-inflated Poisson regression models are used to test which effects are statistically significant.

The fourth and final question relates the variants to their linguistic realisation and thus taps into the second big question presented above: What impacts on the linguistic and multimodal realisation of speech? Previous work has mainly focussed on a lexically-based form function link, without paying much attention to the surrounding talk. However, interactional work has shown that this is precisely where interactional structure impacts on linguistic realisation (Ogden 2006; Nilsson 2015; Raymond 2018). In our analysis of LISTENER RESPONSES and, in fact, any other talk, we need to keep in mind that the linguistic and multimodal realisation of each contribution is formatted with reference to the preceding talk on all levels of linguistic structure, and prosody plays an important role in distinguishing between different actions. This means that we need to consider the lexical, prosodic, and morphosyntactic shape of the preceding segment of talk as a structural constraint on the lexical, prosodic, and morphosyntactic shape of the utterance (here, the LISTENER RESPONSE) under scrutiny. It also alerts us to the necessity of paying close attention to the interplay between prosodic and lexical shape. To this effect, the final analysis chapter (chapter 7) picks up the broad descriptions of the *lexical and prosodic realisation* of the individual ACTION TYPES presented in chapter 5. It shows that the *lexical and prosodic form* of any Listener Response are strongly influenced by the lexical and prosodic shape of the preceding talk. Especially prosodic shape needs to be described relative to the prosody of the preceding segment. It also reminds us that in interaction the prosodic and the lexical cannot be separated, and this needs to be reflected in our analysis. This connection surfaces particularly strongly for LISTENER RESPONSES that are ambiguous based on their lexical form alone, but disambiguated by their prosodic realisation.

In concert, the analysis presented in this thesis demonstrates that LISTENER RESPONSES need to be analysed in the context of the interaction at all possible levels, and how this can be done: from the definition of the variable and the envelope of variation to the description of the lexical and multimodal realisation of each individual variant. *How many* LISTENER RESPONSES get done, *which actions* they do, and *how* they are realised linguistically is strongly influenced by structural constraints rooted in the interaction, here the talk and actions of the main Speaker rather than the person producing the responses.

The remainder of this thesis is structured as follows: the literature review (chapter 2) presents the necessary theoretical and methodological background, elaborating on the gaps and questions both on a theoretical level and with respect to LISTENER RESPONSES and GENDER in particular. The methodology (chapter 3) introduces the data and specific methodology used here. The analysis chapters step through the key challenges in the order presented above, demonstrating that for LISTENER RESPONSES interaction is crucial on all levels of structure, and needs to be integrated into any quantitative analysis. The discussion (chapter 8) ties these findings together and proposes a generalisable model for a sociolinguistic theory and methodology that treats interaction as fundamental, and that can be applied to other discourse-level variables. Furthermore, I discuss how we can also draw on coding for actions or interactional phenomena to better understand sources of phonetic variation.

To summarise, this PhD thesis makes two contributions, one of them theoretical, the other methodological: The contribution to sociolinguistic theory is to show how attending to the organisation of talk-in-interaction improves our understanding of language variation and allows us to tackle the four key questions outlined above. The contribution to sociolinguistic methodology is a clear, reproducible demonstration of how this can be done, in the given case for the discourse-organisation variable (LISTENER RESPONSES), and a first step towards applying this approach to phonological and morpho-syntactic variables.

Chapter 2

Literature Review

> If I have seen further it is by standing on the shoulders of giants.
>
> *Isaac Newton* (1675)

This thesis brings together a number of different literatures and perspectives. Crucially, it will contribute to current debates in variationist sociolinguistics by looking at how we can (a) theorise and analyse variables above the level of the phoneme, and (b) better integrate interactional and variationist approaches on all levels of linguistic structure. It will demonstrate how this is possible based on the example of LISTENER RESPONSES and GENDER.

Therefore, I will first present the current challenges with respect to modelling sociolinguistic variation on and above the level of the phoneme, with particular attention to interactional factors. I will then review relevant Conversation Analytic work, introducing basic concepts, discussing work that shows how closely related interactional function and linguistic realisation are, and reviewing important considerations with respect to CA-based quantification.

Once the theoretical background has been established, I will introduce LISTENER RESPONSES as our case study, with a particular focus on the social variable GENDER. I will first present relevant quantitative, then qualitative, and then mixed methods work. The literature review concludes with a summary of the main theoretical and methodological gaps, and a brief outline of how they will be addressed in this thesis.

2.1 Variationist Sociolinguistics and Interaction

This first half of this Literature Review focuses on building the theoretical foundation and demonstrating gaps at the level of overall theory and methodology. I start from the original definition of the linguistic variable proposed by Labov (1963),

moving on to more recent variationist sociolinguistic studies that try to bring to-gether phonetic variation with aspects relevant to the interaction these variants occur in. I then discuss the challenges scholars who study linguistic variables above the level of the phoneme encounter and argue that these come about because the definition of the variable and the variants are not rooted in (interactional) structure. Conversation Analysis – the focus of the second half of the theoretical Literature Review – appears a natural ally if we want to address these challenges.

2.1.1 Conceptualising the sociolinguistic variable

In the early 1960s, William Labov laid the foundations for variationist sociolin-guistics as we know it today. His definition of the sociophonetic variable is the foundation of a fruitful research tradition that has grown to include aspects of the interaction in which these phonetic variants occur, as well as variables above the level of the phoneme. Nevertheless, there still is tension between the traditional Labovian definition of the variable, and more interaction-focussed approaches.

2.1.1.1 Sociophonetic variation

It is essential to understand the foundations and traditional definition of the linguistic – in this case phonological – variable before critiquing the more recent definitions of variables above the level of the phoneme. The key points of critique are that neither Labov's definition nor those of variables above the level of the phoneme take into account interactional aspects, and that they cannot necessarily be transferred to variables at other levels.

   To put it very generally, a variable is a thing that can be done in different ways. Accordingly, two ways of doing the same thing are variants of how this thing can be done. In his pioneering study of linguistic variation on the island of Martha's Vineyard, Labov (1963: 279) focuses on phonological variables, particularly vowels, and lays out the four 'most useful properties of a linguistic variable': (1) it should be *frequent*, (2) it should be *structural*, or 'integrated into a larger system of functional units' as much as possible, to make it interesting with respect to linguistic variation. (3) Its distribution 'should be *highly stratified*', i.e. vary between social groups in some obvious manner, to make it interesting with respect to social variation. And (4) two additional contradictory criteria need to be considered: (a) the variable should be *salient* to both speakers and observers, because this allows us to also look at relationship between social attitudes and linguistic behaviour, but (b) at the same time it should be one speakers do not consciously manipulate, in order to reduce the observer's paradox.

After defining these four key properties of a linguistic variable, Labov proceeds to defining the variables and their variants, to analysing the linguistic constraints, then the social constraints, and then the interaction between the linguistic and the social. Six decades of work, including this PhD thesis, are modelled on this structure.

This original definition and the vast majority of variationist work following in Labov's footsteps focus on phonological variables. However, the criterion that a linguistic variable should be 'structural', so that it can be analysed with respect to constraints internal to the linguistic system, *can* be transferred to discourse-level work. This could help resolve a tension that is only one decade younger than the Martha's Vineyard study.[1] There is a wide variety of sociophonetic work that takes into account not only social factors like age, gender, or socio-economic class, but also group identities that are locally constructed (ethnographic work like for example Eckert (1989) and Mendoza-Denton (2008)), or even more local meaning-making processes like constructing a specific persona (Podesva 2007; Mendoza-Denton 2008; Hall-Lew et al. 2017), stance-taking, or style-shifting for other reasons (Holmes-Elliott and Levon 2017; Moore 2004; Barnes 2018).

However, within the quantitative paradigm there is very little work that takes into account how the intrinsically interactional nature of communication impacts on sociophonetic variation. The next subsection will review the first steps towards integrating interactional and variationist analyses, primarily at the level of phonological variation.

2.1.1.2   ... and interaction

In the structural definition of the sociophonetic variable, interaction is irrelevant – the variable is a sound that can be realised in different ways, without changing the meaning of the word it occurs in (semantic equivalence). However, with respect to the broadly social factors influencing the realisation, scholars have moved from macrosocial categories mentioned above to increasingly more locally relevant categories. But even work on stance and style or persona construction does not engage with the interactional nature of language and the structure underlying language production in interaction as deeply as it might.

A few recent studies show how interactional or Conversation Analysis can be linked to analyses of language variation at the phonological, but also morphosyntactic and lexical level. Chakrani (2015) and Raymond (2018) stay at the qualitative level but demonstrate that participants in an interaction strategically

---

[1] I will discuss the challenges and debates that have ensued with respect to variables above the level of the phoneme (Sankoff 1973; Lavandera 1978; Dines 1980) after a brief note on interaction.

use language variation and orient to different variants in their talk: Raymond (2018) analyses radio phone-ins in Spanish-speaking radio in the USA and notes that callers (inadvertently) position themselves as speakers of different varieties of Spanish through their use of specific lexical, phonetic, and morphosyntactic variants, and hosts sometimes comment on these or playfully recycle them. This paper demonstrates that language variation is an important resource in interaction. Similarly, Chakrani (2015) analyses interactions between friends who speak different varieties of Arabic and pick up on each other's linguistic features, often to humorous effect.

Nilsson (2015) analyses Swedish talk-in-interaction with respect to several dialect variables that are levelling. She finds that speakers who generally use the traditional dialectal variant draw on the levelled variant when format-tying with a person who uses the levelled one more frequently, and vice versa. This suggests the choice of a phonetic variant is tied to what participants are *doing* in the interaction, and this can be related back to overall quantitative patterns of variation. We will see later on in the review of Conversation Analytic work that CA scholars have long known that we design our talk specifically to suit the interactional context it stands in on all levels of linguistic and multimodal structure – but CA traditionally has no interest in quantifying this and comparing patterns of variation across groups, social or otherwise.[2]

This very brief review shows that it is both relevant and possible to bring an understanding of interactional structure into the analysis of phonological language variation as a broadly social (or interactional) constraint. Language, of course, also varies with respect to variables above the level of the phoneme, and particularly for variables at the discourse-level interaction can and in fact should be considered a structural constraint. The next section of the literature review will focus on work on this type of variable, and the challenges of not taking into account the interactional nature of language in our variable definition and analysis.

2.1.1.3 Variables above the level of the phoneme

Only about a decade after Labov's first studies of phonological variation on Martha's Vineyard and in New York, scholars began to extend the notion of the variable to other levels of linguistic structure, including syntax (Sankoff 1973) and what has come to be called discourse markers or discourse-pragmatic variables (Dines 1980;

---

[2]'Talk in interaction is about constructing actions, which is why it does not reduce to language; treating talk in inter-action only for its properties as a system of symbols or a medium for articulation or deploying propositions does not get at its core. And the actions that are constructed by talk and other conduct in interaction compose, and are parts of, trajectories or courses of action, which is why a pragmatics that does not attend to the sequential organization of actions is at risk for aridity.' (Schegloff 2015: 355)

Lavandera 1978). Particularly work at the discourse-level encounters three key challenges:

1. how to define the variable and the variants in a 'structural' way, so that

   (a) these structural constraints can be analysed, and

   (b) the principle of accountability is observed (ie all *possible* occurrences are noted, not only all those observed),

2. how to quantitatively analyse the variable's frequency of occurrence, and the distribution of the variants, and

3. how to deal with the multifunctionality of individual items.

The definition of the sociolinguistic variable above the level of the phoneme has long been an issue of contention. Broadly speaking, variants of one variable are different ways of doing or saying 'the same thing'. What this 'same thing' is, however, depends on the variable definition, and that in turn depends on the kind of question we are asking. With respect to syntactic and discourse features, semantic, functional, formal, or derivational equivalence have been proposed and applied as definitional criteria (Buchstaller 2006a; Buchstaller 2008; Lavandera 1978; Dines 1980; Pichler 2010; Pichler 2013b; Tagliamonte 2016; D'Arcy 2017). Buchstaller (2009) presents an excellent summary of the underlying issues and approaches taken to defining and quantifying morphosyntactic variables. Waters (2016), seven years later, notes that given the diversity of discourse-pragmatic variables, each needs a 'bespoke analysis'. One challenge common to all of these approaches is satisfying the principle of accountability, and creating unambiguous coding schemes. Both of these concerns can be addressed when conceptualising LISTENER RESPONSES as a discourse-organisational variable.

Situating the definition of the variable and variants in interactional structure builds on Schiffrin's discourse approach and Maschler's interactional grammar perspective on language and gender (summarised in the excellent overview in Maschler and Schiffrin (2015)). Neither of them have received much attention in Discourse-Pragmatic work so far.

Some variables, like quotatives, can be defined and delimited in a way that makes it possible to also count non-occurrences: if we functionally define them as 'all strategies used to introduce reported speech, sounds, gesture and thought by self or other' Buchstaller (2006a: 5) it becomes possible to 'close the set' and

satisfy the principle of accountability by also counting the absence of quotatives in contexts where we might expect one.[3]

For other discourse-pragmatic variables like Like (D'Arcy 2017) or Neg-tags (Pichler 2013b) it is very difficult to satisfy the principle of accountability and count not only presences but also relevant absences of the variable (they are thus what we would consider 'open set' variables). This, in turn, means the variable's frequency needs to be normalised with respect to some other feature. This has often been done relative to the overall size of the corpus and expressed as occurrences per million words, or some similar metric. However, these approaches completely remove the variable from its interactional context and do not allow for any consideration of potential structural constraints. This can lead to gross misrepresentations of what actually happened in the interaction, as Murphy (2012) discusses with respect to Listener Responses and gender. I summarise her argument in the second half of the literature review focussing on this particular variable.

Additionally, these frequency counts make it difficult to run inferential statistics on the variable distribution that take into account structural factors in a cohesive manner. Discourse-Pragmatic Variation uses varbrul as a tool which does make it possible to compare frequencies and derive statistically (non) significant differences between groups, but the underlying problem remains: it is difficult to analyse structural constraints and include them when analysing the social constraints if the variable is not rooted in 'structure' in some explicit manner. This might then lead to seemingly social differences that in fact are due to un-theorised structural constraints.

This leads us to the second big issue mentioned above: both Like and Neg-tags can serve numerous different functions, which need to be coded in some way. Because the variable itself is not rooted in interactional structure, the variants cannot be either. This leads to three complications. The first relates to the emic validity and the comparability of categories and findings across studies: Researcher-imposed definitions of variants are not necessarily locally relevant to the interactants, and can be difficult to agree on between coders. Inter-coder reliability is a common tool to assess how intersubjectively reproducible coding categories are. Would it not be even better, though, to have a way of developing coding categories that are based on the participants' orientations in the interaction, that are developed bottom-up from the data and thus emically valid? This could also address the second complication with top-down definitions of variants: they

---

[3]We can think of variables where we can circumscribe the envelope of variation in a way that allows us to also count absences as 'closed set' variables, and those where it is not possible to do this as 'open set' variables. I will elaborate on this difference in the methodology section.

are often rooted at different levels of structure, mixing for example pragmatics, semantics, notions of alignment and affiliation, preference, or politeness. This, in turn, leads to a third complication: if the variants are coded on several levels simultaneously, they become seemingly multifunctional, which complicates our analysis further. If one token has several functions, we cannot even calculate percentage distributions of the frequency of those functions – they will not sum up to 100 %. Again, defining functional variants in interactional structure could address this, because it clearly defines one layer of function under scrutiny. Other and further layers can be added on, but do not conflict with each other. Rather, they would need to be analysed individually first, and then in interaction with each other.

In their analysis of tag questions, Moore and Podesva (2009) code for topics of conversation as well as stances taken, which introduces some functional and interaction-based differentiation, but they do not really delve into the interactional structure. Similarly, Buchstaller (2011) alludes to the importance of the larger interactional project participants embark on when it comes to quotatives, but given that this is only a small section of her overall analysis does not go into more detail about the implications. At the more local level, Buchstaller (2008) and Buchstaller (2014) points out the manifold epistemic and stance-related aspects related to quotation. Turning this into a fully interaction-based coding scheme would be a large and challenging project.

This brief theoretical review shows challenges of defining particularly 'open set' variables and their variants above the level of the phoneme in a way that is clearly rooted in a structure, which then would allow us to (1) analyse structural constraints, and (2) quantify both overall frequency of occurrence and the distribution of the variables descriptively and inferentially. This lack of a structural definition also leads to complications with respect to multifunctionality.

Pichler (2013a) explicitly calls for employing Conversation Analytic tools and methods in our analyses of open set discourse-pragmatic variables to address these challenges. She introduces the ideas of recipient design, preference organisation, turn allocation, topic management, sequential implicativeness, adjacency pairs, and the next-turn proof procedure, and draws on different sets thereof in the different analysis chapters.

While she lays important groundwork both in this chapter and in the book overall, Pichler (2013b) does not turn this 'Conversation Analytic orientation' into a cohesive theoretical and methodological approach ready to be transferred to other variables. Furthermore, the notion of 'preference organisation' can be difficult to take out of the CA context, because of its very specific meaning in this analytic tradition. 'Preference' in CA refers to how things are 'normally' done, and what

is 'expected' as a next action, but can only be inferred based on the *orientations* participants display – an action can be marked as dispreferred in various ways, for example by hedging, using fillers, openly apologising, or first doing a pro forma preferred action, followed by the dispreferred one (see for example Pomerantz (1984)). Outside of a CA context, 'preference' frequently is treated as a psychological notion of participants 'liking' one option better than the other. Pichler (2013a) also mixes the local and the overall level of structural organisation (more on this in the next section). These two levels of course go hand in hand and impact on each other. However, similar to the issue of multifunctionality just discussed, it is analytically more clear to keep them apart in the first analysis step, and then recombine them.

All these challenges are issues which surface once we begin to integrate Conversation Analytic or broadly interactional approaches into our analysis of variation, and important stepping stones for developing a more cohesive theory that draws on both traditions. I build on Pichler's work and show how this could be done, and how far a full integration of interactional and variationist methods can take us. In order to do so, let us now briefly review some Conversation Analytic concepts and relevant work, and then consider how we can fruitfully integrate these two approaches.

2.1.2   Variation and Conversation Analysis

Where work on language variation at all levels so far does not take into account the effect of interactional structure, Conversation Analysis in its conservative form objects entirely to quantifying (see Schegloff (1993), Steensig and Heinemann (2015), Nishizaka (2015), and Ruiter and Albert (2017)). On the other hand, there are several traditions of Conversation Analytic work that show links between linguistic form and interactional structure or function, and while Schegloff (1993) argues that it is not (yet) possible to quantify in a way that is true to CA, he lays out very clear criteria for CA-based quantitative analysis. Only six years later, Heritage (1999) says that we do know enough to do it, albeit always with caution.

Before delving into CA work describing linguistic variation without quantifying, and the discussion about quantification and CA, let me briefly introduce the Conversation Analytic approach to interaction. Conversation Analysis is the study of the *structure of human social interaction* (Schegloff and Sacks 1973; Jefferson 1978; Sacks et al. 1974; Atkinson and Heritage 1984), with a special focus on how we use language in this context. It has its intellectual roots in ethnomethodology and

Goffmanian sociology.[4] CA is interested in the ordinariness of human interaction, in the daily achievement of smooth speaker-transitions and successful interactions. In order to understand the structures governing these interactions, CA scholars work with large *collections* of 'similar cases'; they study one phenomenon, for example assessment sequences (Pomerantz 1984), based on close analyses of many instances from different social and interactional contexts.

Conversation Analysts differentiate between the overall structural organisation (see Robinson (2012)) and the local level of organisation. The overall organisation refers to how interactants move from one project to the next, for example from topic to topic, or from an activity like 'sharing news' to 'planning a visit'. The local level of organisation is focused on who says what, and when. It includes sequence-organisation, turn-taking organisation, and topic management on a more local level (see Stivers (2012)). I focus on this local level of organisation here, in order to develop my theoretical and methodological contributions.

Whenever any party in an interaction says or does anything, the Conversation Analyst asks 'Why this now?' (Schegloff and Sacks 1973: 299). The meaning or interactional function of any contribution in an interaction is determined not based on the form of that contribution, but on how the person it was addressed to treats it in the next turn. This so called 'next-turn proof-procedure' is one of the core tools in Conversation Analysis.

The most important notions for us to briefly touch on are the turn-taking organisation and the adjacency pair, and building on this the just mentioned notion of the next-turn proof-procedure:[5] Humans take turns at doing things, and one thing always builds on the other. I ask you a question, you give me an answer. You greet me, I greet you back. I invite you, you accept or decline – and usually give a reason for declining, or preface and soften a rejection in some other way. These action pairs are considered adjacency pairs – there is a 'first' and a 'second', where the 'second' is the reaction to the 'first'. Importantly, what the 'first' does or means is negotiated in the interaction. Now imagine I ask what might be a question, you do not answer it, and I do not follow up on the question, then we have agreed in our interaction that what I have done was not, in fact, ask a question. Alternatively, if I ask a question, you do not answer it, and I ask it again or draw your attention to the need to answer in some other way, I am orienting to your lack of response

---

[4]For a more in-depth review of the origins and development of what has come to be known as 'Conversation Analysis', see Maynard (2012). Hoey and Kendrick (2017) further provide a concise introduction to CA for psycholinguists that is also helpful to more quantitatively orientated sociolinguists.

[5]For more in-depth explanations and introductions to CA see for example Clift (2016) and Sidnell and Stivers (2012). I also highly recommend Goffman's exceedingly well-written 'Interaction Ritual' (Goffman 1982) and 'Forms of talk' (Goffman 1981) as introductions to some of the intellectual roots of the field.

as a problem and explicitly mark my question as one. This way of understanding any action based on the re-action it elicits is the next-turn proof-procedure. It will be described in more detail in chapters 4 and 6, where we will see it in action and see its implications for our definition of the variable and the variable context.

2.1.2.1 Non-quantitative CA work on variation

As mentioned, several strands of CA work show that variation in the linguistic realisation of an utterance is systematically related to the interactional work that is being done in this utterance (or the way this utterance is treated in the next turn). However, within CA, quantitative analyses of variation are not deemed relevant, and within studies of language variation, this Conversation Analytical awareness that every contribution's exact format is related to its interactional function has gone largely unnoticed.

Existing sociolinguistic work which is attentive to the interaction shows that the local organisation of interaction has an influence on the phonetic and prosodic shape of utterances (see for example Barth-Weingarten et al. (2010), Selting and Couper-Kuhlen (2004), and Couper-Kuhlen and Selting (1996)). Similarly, participants have been argued to show a 'prosodic orientation' (Szczepek Reed 2007; Szczepek Reed 2009; Szczepek Reed 2010; Szczepek Reed 2011) by adjusting their speech rhythm and the prosodic shape of their utterances to signal whether they are formulating a list (Selting 2007),[6] and use phonetic features to signal whether they are starting a new action or continuing the action from the previous segment of talk (Szczepek Reed 2014).

Similarly, work on the phonetics of talk-in-interaction (Ogden 2012) shows that the exact phonetic design of the 'edges' of utterances gives interactants essential information about whether a contribution ends an ongoing turn, or projects further talk by the same speaker (Local 2003; Local and Walker 2008). Furthermore, studies of preference organisation and other interactional phenomena have shown that these activities situated at the local level influence lexical and morpho-syntactic choices (Pomerantz 1984; Ogden 2006), as do processes like format-tying (Nilsson 2015) or alignment and affiliation (Gorisch et al. 2012; Stivers 2008).

Conversation Analytic and interactionally oriented scholars focussing on the relationship between intonational patterns and functions have further investigated this form-function relationship for small tokens like for example JA, JAJA (Golato and Fagyal 2008; Barth-Weingarten 2011), or ACHJA (Betz and Golato 2008), as well as UH-HUH and YEAH (Schegloff 1982; Drummond and Hopper 1993a). Each of

---

[6]This is hence situated at the level of the overall structural organisation.

these can fulfil a number of different interactional functions, depending on (a) its sequential placement, and (b) its prosodic shape.

Barth-Weingarten (2011) demonstrates that the German response token JAJA can have five different interactional functions, depending on its phonetic and prosodic realisation – including pitch contours as well as smiley voice – and multimodal cues like nodding. These range from (re)claiming epistemic authority to aligning with a joke. She notes that it is important to consider sequential placement, sequence-organisational function, interactional function, and sequential consequence in the analysis. This explicitly separates the 'interactional' from the sequential-structural and offers us a way of resolving the multifunctionality scholars focusing on variation at the discourse level from a quantitative point of view have long been struggled with. The same token does fulfil different functions, but these are situated at different levels of structure. Hence, the trouble with multifunctional tokens might be resolved by carefully separating those levels of structure in the analysis, and then paying close attention to co-occurrence patterns.

Related work on YEAH and MHM is discussed in the section on qualitative work on Listener Responses; the variable I will use to exemplify the theoretical and methodological challenges outlined at a more abstract level in this first half of the Literature Review. This large and diverse body of interactional work clearly suggests that interactional structure impacts on the linguistic realisation of each utterance, from phonology to prosody, and that prosodic and lexical shape are inextricably intertwined. However, this knowledge or even just an awareness of the importance of interactional structure is only very rarely integrated in our analyses of language variation.

2.1.2.2 CA and Quantification

It might seem surprising that this knowledge is not translated into quantifiable coding schemes or approaches to language variation – certainly the local level of organization must matter, given that the shape of the language we produce all the time in interaction is clearly impacted by the things we *do* in these interactions. However, within Conversation Analysis itself, quantification is simply not a relevant thing to do. Instead, CA traditionally aims to uncover the *structural patterns* underlying (all) human interaction, as distinct from variation in how exactly different groups of humans apply and modify these patterns or rules (Sacks et al. 1974).

At the same time, Conversation Analysis is no stranger to 'informal quantification', drawing on notions like 'many' as opposed to only 'a few' cases in which we can see a certain pattern. Further, since the (extremely cautionary) discussion of the sense(lessness) of quantification in CA by Schegloff (1993), an – albeit small

– strand of research has developed in which quantification is being done. For (traditional) Conversation Analysis itself, quantification might not seem particularly relevant, though several scholars argue there is something to be gained:

Steensig and Heinemann (2015) point out that first of all, using quantitative coding in CA forces the scholar developing and using this coding scheme to very rigorously question their category assignments: How do I define the phenomenon I am analysing? What are typical, what marginal cases? Why? How do I chose the cases that go into my collection? Secondly, by developing such a coding scheme and potentially quantifying some of the findings, previously hidden gaps in our understanding of the structural organization of human social interaction, and avenues for future research can be uncovered. Thirdly, and more generally, Kendrick (2017) argues that introducing quantification into CA is a sign of the field maturing – a development parallel to that of sciences like biology or physics after their inception. One example of how quantification can move forward CA theory is a study published by Zama and Robinson (2016) looking at long and short blinks as response actions in narrative environments. They and Ruiter and Albert (2017) cite further studies where quantification helped drive CA theory forward.

Furthermore, being open to quantification and developing coding schemes based on interactional structure opens up exciting opportunities for cross-disciplinary collaboration like the one presented here (see Stivers (2015), Ruiter and Albert (2017), and Kendrick (2017) for a more extensive discussion). Schegloff (1993) himself critiques the analysis of *Oh* in response to questions in a sociolinguistic interview presented in Schiffrin (1987a). He argues that in order to qualify her finding that *Oh* usually prefaces responses to unexpected or surprising answers, Schiffrin (1987a) would need to do a CA-based qualitative analysis that takes into account the participants' orientation to the situation and the interactional import of the *Oh* in theses responses, following Heritage (1984) (whose findings are not entirely aligned with hers).

Despite this practical example of how CA-based quantification could be useful, Schegloff (1993) claims that it is simply not possible at the time of writing. Only a few years later, Heritage (1999) argues that by the end of the 20th century we do have enough CA work on specific practices that allows us to quantify in a manner that is accountable to the interaction. He points out that interactional practices as variables in these studies could be construed as dependent variables, i.e. something we would expect to change based on participants' identities or the context, but also as independent variables, i.e. something that influences outcomes and shapes the context of interactions and participants' identities. This is precisely what I will be doing in this thesis, and I will explain what the different variables and variants are at the start of each analysis chapter. Since Heritage (1999),

quantification has increasingly been used successfully in CA studies, including collaborations with different disciplines (see special issue of ROLSI from 2017).

2.1.2.3 How to quantify in an interactionally accountable way?

Given that quantification in CA is possible, has been and is being done successfully, and comes with potential benefits like inter-disciplinary cross-fertilisation and developing new questions for CA itself, we need to ask: What does a successful, responsible CA-based quantitative analysis look like? Steensig and Heinemann (2015) and even more so Schegloff (1993) outline very clear guidelines along which a quantitative analysis of an interactional practice can be developed.

First of all, to use Schegloff's terms, we need to develop a clear definition of the denominator, the numerator, and the context (importantly, to him what the context is depends entirely on the orientations the participants display, not on any external factors). The denominator and the numerator are what Stivers (2015) refers to as a characterisation of the phenomenon in question with respect to its sequential position and its linguistic and multi-modal composition, as well as a clear specification of all sub-types of the phenomenon.

In variationist terms, the denominator refers to the variable and the envelope of variation, and the numerator to the variants (including null-realisations where relevant). If we aim to compare behaviours between different groups of people or different languages, we need to distinguish between groups or contexts, and here lies one important difference already alluded to above: for Schegloff (1993) these groupings need to be grounded in and developed bottom-up based on the participants' orientations, not superimposed as top-down macro social categories.

It is important, of course, to keep in mind that (1) every coding scheme is knowledge frozen in time, and (2) creating clear cut-off points between categories necessarily flattens the complexity of real interactions (Stivers 2015), and that we need to carefully reflect on our theoretical and methodological premises (see (Nishizaka 2015)). This means reflecting on and being clear about what we define as the variable and the variants, what we consider evidence, and what status 'numbers' have in our analysis. If these concerns are adequately addressed, a combination of CA and quantitative methods is possible and an exciting way forward that might allow us to bridge the gap between CA and quantitative studies of language variation.

2.1.3 So where is the link?

Let us review the main challenges in work on language variation with respect to accounting for the interactional nature of language, and then outline how

Conversation Analytic tools can be useful to address these gaps. The key challenge relates to analysing variables above the level of the phoneme, particularly the ones referred to as discourse markers or discourse-pragmatic variables. It can be broken down into four smaller challenges: (1) defining and delimiting the variable and the envelope of variation, and relatedly quantifying its frequency in an accountable way, (2) defining the variants and (3) analysing their frequency distribution, taking into account structural constraints, and (4) understanding the link between the linguistic form and the interactional function of the individual variants. Answering the fourth question will then allow us to move from discourse-pragmatic to phonological and other linguistic variables and better understand how interactional and phonological structure interact.

The challenges outlined above come about because discourse-pragmatic variables are generally not defined in a way that includes one of the 'most useful properties' of linguistic variables (Labov 1963) – that of being 'structural'. An interactional, Conversation Analytic approach taking into account the requirements outlined by Schegloff (1993) can address this, especially considering that CA investigates the *structure* underlying human social interaction: if we define the variable as rooted in interactional, sequential structure, the variants become those things that can happen in this sequential slot. They need to be defined and described based on a close analysis of the interaction, relying on the participant's orientations (next-turn proof-procedure). The precise ways of accountably quantifying will need to be explored for each variable individually (see Waters (2016) on the need for a 'bespoke analysis'). There is already some initial work on the link between interactional function and linguistic form (Nilsson 2015; Chakrani 2015; Raymond 2018).

An analysis of one discourse-level variable stepping through the four questions outlined above can show in more detail how the exact linguistic realisation is linked to what participants are doing in the interaction. Thus, all in all, integrating Conversation Analytic tools into our analysis of variation allow us to address these gaps, provided we choose a variable that we already have a good qualitative understanding of, or – for less well-explored variables – a lot of time and resources to develop this understanding.

Now that we have established the theoretical questions in variationist sociolinguistics and the necessary background on tools and considerations in Conversation Analysis, and seen some first examples of how these two approaches have been successfully brought together, let us zero in on the case study under analysis in this thesis: LISTENER RESPONSES as a discourse-organisational variable.

2.2 Listener Responses as a discourse-organisational variable

What I call LISTENER RESPONSES here has been introduced to the quantitative literature as *backchannels* (Yngve 1970) and is still frequently referred to in this way (Oreström 1983; White 1989; Wong and Kruger 2018; Kogure 2003). Other scholars refer to the phenomenon for example as *minimal responses* (Fellegy 1995; Reid 1995), *reactive tokens* (Clancy et al. 1996), or *response tokens* (O'Keeffe and Adolphs 2008).

This phenomenon is an ideal case study to develop a theory and methodology that integrate interactional and quantitative variationist methods for three main reasons: First of all, even the traditional definition of LISTENER RESPONSES clearly characterises them as rooted in the interaction – but does not follow through on the implications of this in the analysis. Second, there is ample qualitative and quantitative, as well as some mixed methods work on them, and third, this work exemplifies all the challenges presented in the theoretical part of the literature review. This means they can be considered an established variable – with the challenges that come with defining variables above the level of the phoneme – and there is enough interactional work to fulfil the requirements set out by Schegloff (1993) and Heritage (1999) that we need to understand the qualitative basis of the variable well enough before proceeding to any sort of quantification.

Most quantitative work on LISTENER RESPONSES focuses on either culture/L1 or gender as social variables. Because gender is better represented in my corpus I will primarily focus on these papers here. However, the same challenges apply to all social variables that have been analysed. The core issues I will problematise in this literature review, and then propose solutions to, are as follows:

1. Past work on LISTENER RESPONSES has looked at vastly different subsets of the phenomenon, which makes cross-study comparison challenging, and a quantitative understanding of the phenomenon as a whole even more difficult. We need not only a common denominator, but also a common demarcation.

2. Scholars have used numerous different approaches to quantifying LISTENER RESPONSES and to normalising their frequency of occurrence, none of them sensitive to the interaction. This leads to similar practical problems as above.

3. The core critique of the original work on backchannels (Yngve 1970) raised by Conversation Analysts (Schegloff 1982) is that a very diverse set of actions is treated as one. Thus, it becomes essential to develop interactionally rooted

sub-categories that then also allow us to take into account the structural constraints on variation between these categories. So far, only a few 'mixed methods' studies have used function-based descriptions of the individual variants. However, these are not necessarily rooted in interactional structure. Relatedly, to my knowledge there currently is no inferential statistical method that would allow us to analyse between-group variation; so far this has only been done based on percentage distributions.

4. There has been debate as to which tokens fulfil which functions, notably with respect to *yeah, uh-huh*, and *mhm* (Schegloff 1982; Jefferson 1984a). We should revisit the link between form and function for LISTENER RESPONSES.

In the following sections, I will discuss quantitative, qualitative, and mixed methods work to illustrate the points raised here, with a particular focus on studies that treat gender as their social variable. I will then summarise the key questions and gaps presented, and outline how this dissertation addresses them in the following chapters.

2.2.1 Quantitative work on Listener Responses

Here, I will focus on work that only looks at *how often* Listeners respond, without differentiating *what* is actually being done with these responses. The section on qualitative work will then focus on different actions Listeners can do, and the third section on mixed methods work reviews previous approaches to taking the different functions or types of LISTENER RESPONSES into account.

There are two big gaps with respect to overall LISTENER RESPONSE frequency, equivalent to the first two of the points raised above. First of all, we need a definition based on their sequential and interactional structure. At the moment, there is a common denominator as 'things Listeners do', but no common demarcation that would allow us to frame studies as looking at a particular subset. This leads to a lack of comparability across studies, and makes it even more challenging to construct a big picture including all the different studies. Second, LISTENER RESPONSE frequency has been operationalised in very different ways, most of which are not sensitive to the interlocutor. This, too, makes comparability across studies rather difficult and leads to different and contradictory findings with respect to variation by gender or culture.

2.2.1.1 Defining and delimiting the phenomenon

In past quantitative work, LISTENER RESPONSES have been defined in terms of their placement and duration (Schweitzer and Lewandowski 2012), based on a specific

set of lexical items (Fellegy 1995; Gardner 1998) or interactionally and sequentially (Gorisch et al. 2012; Norrick 2012).[7]

Schweitzer and Lewandowski (2012) take a fully automated approach and define as a backchannel any utterance shorter than one second in between utterances of the other speaker. The advantage of automation is not to be denied, but this definition is highly restrictive and potentially misses a good number of LISTENER RESPONSES, given that (a) they can easily be longer than one second, (b) they can be in overlap with the other speaker's talk, rather than only occur in pauses between turns, and (c) multimodal cues are not even mentioned in this study.

The majority of scholars in the quantitative paradigm take a practical approach to the phenomenon, restricting their analysis to a small set of tokens, usually including forms of *mhm, uh-huh, yes, yeah, mm* that make up about 75 % of everything considered a LISTENER RESPONSE in the data set under scrutiny. White (1989) and Kjellmer (2009) explicitly give these numbers, while others simply define listener responses as a finite set of tokens (Reid 1995; Fellegy 1995; Heldner et al. 2010; Cathcart et al. 2003), for example all instances of *yeah, mhm, uh-huh* and *right*. They usually do gesture at the relevance of multimodal cues and the existence of different and longer LISTENER RESPONSES, but decide to exclude them from analysis for practical reasons.

Some scholars focus on multimodal cues; Brunner (1979) for example shows that smiles can be used as backchannels, and Maynard (1990) includes head nodding and laughter in her analysis. Bavelas et al. (2000) and Kogure (2003) include all of the above, as well as non-minimal responses. Most quantitative papers only gesture towards the existence of non-minimal responses but refrain from analysing them. This is probably because the boundaries between a LISTENER RESPONSE and a turn or a grab for the floor are more difficult to determine the longer the responses are. Ward (2006) even claims that the notions of *turn* and *floor* are far too problematic to use for analysis. Accordingly, it is more practical to restrict oneself to clear-cut cases of fairly minimal responses. Some scholars accordingly opt for a narrow definition of LISTENER RESPONSES that relies on their brevity.

This brief overview shows that what is considered and analysed as a LISTENER RESPONSE varies greatly: some scholars use a definition based on form, others base it on placement. Function, however, is not treated as a definitional criterion. Different research questions certainly warrant analysis of a different subset of

---

[7]This interactional and sequential definition is common in the qualitative work on LISTENER RESPONSES, (Goodwin 1984; Jefferson 1984a; Schegloff 1982). I will introduce these in the next subsection.

LISTENER RESPONSES, but it would be extremely useful to have one clear, shared definition that is not only based on a common denominator but considers the set of LISTENER RESPONSES under scrutiny a *subset* of the whole set of LISTENER RESPONSES.

The restrictive definitions are clearly useful to answer certain kinds of research questions, but also miss at least a quarter of the things that can be considered LISTENER RESPONSES. To fully understand how this phenomenon patterns, we need to model the whole set, not only three quarters of it.

2.2.1.2 Operationalising Frequency

There are four main approaches to quantifying the overall frequency of LISTENER RESPONSES in the literature, most of which do not pay attention to the interactional structure in which the variable occurs. On the one hand, this means findings are not comparable across studies. On the other hand and on a much more fundamental level, it also means these studies misrepresent what is happening in the interaction. Murphy (2012) illustrates that not quantifying based on interactionally sensitive criteria can lead to skewed results. She finds gender differences in frequency that are actually due to the distribution of Speaker- and Listener-roles in the interaction. I would like to suggest that we need to quantify LISTENER RESPONSE frequency relative to the talk they are responding to. This builds on Murphy's (2012) points and a model of quantifying *backchannel* frequency used in social psychology (Duncan and Fiske 1977) that seems to have been overlooked by the quantitative enquiries in linguistics.

Let us first take a brief look at the different ways scholars have quantified and normalised LISTENER RESPONSE frequency; I will draw on this in the first analysis chapter. Corpus Linguists like O'Keeffe and Adolphs (2008) and Murphy (2012) traditionally normalise frequency as 'number of occurrences per one million words'. This statistic is useful to get an idea of how frequent which *backchannel* token is in a given variety of English (or a given language), but completely insensitive to context, interlocutor, opportunities to produce a LISTENER RESPONSE, and obscures both inter- and intra-speaker variation. Hence the usefulness of this type of statistic depends entirely on the research questions asked.

Brunner (1979) and White (1989) orient more strongly to Labov's 'principle of accountability' (Labov 1972: 72), counting what they consider opportunities to produce a LISTENER RESPONSE and then relating these to the number of responses observed. However, the two studies use different definitions of these opportunity spaces, and they find that both verbal and multimodal LISTENER RESPONSES can occur both in overlap with talk and in pauses between TCUs, and that they are

largely optional (their continued withholding, however, would be accountable). This makes it impossible to really quantify the number of possible occurrences, rendering this statistic rather problematic.

Clancy et al. (1996: 368) compare what they call *reactive tokens* across corpora based on the number of reactive token speaker-changes relative to all speaker changes, thereby creating a ratio of how many speaker changes are 'supportive', and how many correspond to actual changes of the floor. This is a very interesting strategy to circumvent the questions of where a LISTENER RESPONSE might be 'relevant' (and also 'relevantly missing'), and could be easily automated. In the presented form, it does not give participant-specific numbers, but this could be adapted by creating a ratio for participant A and a separate one for participant B in each interaction.

Other scholars normalise the frequency of LISTENER RESPONSES relative to time; per minute (Tottie 1991; Bavelas et al. 2000), or per three minutes (Maynard 1990). Oreström (1983: 130) finds that in turns longer than 30 seconds, usually 4-10 seconds pass between LISTENER RESPONSES. These numbers are not as meaningful as they could be, because they do not acknowledge the turn-holder's behaviour – when a participant is in the Listener role, they have more opportunities to produce LISTENER RESPONSES than when being in the Speaker role. We would thus need a metric that at least normalises the number of LISTENER RESPONSES relative to interlocutor speaking time, as in Schweitzer and Lewandowski (2012). Given that speech rates vary and one participant might produce much more talk in one minute than the other, an even more useful metric would be how many words a turn-holder utters between the Listener's responses. White (1989: 63) provides this sort of number – she shows that Japanese Listeners produce one *backchannel* for every 14 words, while American Listeners only produce one backchannel for every 37 words. Unfortunately she does not make explicit whether this is (a) one *backchannel* per 14 interlocutor words, (b) one *backchannel* per 14 words the person producing the backchannels utters, or (c) one *backchannel* per 14 of all the words produced jointly in a given conversation.

As mentioned above, Murphy (2012) illustrates the theoretical and method-ological challenges of the quantifications summarised so far. She analyses 're-sponse tokens' based on age and gender. She draws on CAG-IE, the Corpus of Age and Gender for Irish English, a corpus of self-recorded interactions between friends, stratified according to age (20s, 40s, 70/80 years) and gender (m/f), with roughly 15 000 words per cell. Murphy (2012) defines response tokens based on lexical form and includes 21 single-word tokens. She gives both raw frequencies in the corpus and frequency per million word for each individual minimal response.

Overall, she finds that men across all age groups have a higher frequency of response tokens per million words in this corpus than women do. Men produce about 34000 responses per million words, which means they utter on average 3.4 responses per 100 words, while for women the frequency varies across age groups from 2 to 3 minimal responses per 100 words Murphy (2012: 336). She notes large between-speaker variation and finds that those speakers with very high response token frequencies per million words are those whose interlocutors do the most talking – or in other words, Listeners produce more response tokens than Speakers.

Murphy (2012: 345) thus concludes that it is not necessarily *gender* that correlates with response token frequency, but that speaker roles, relationships, and other contextual factors need to be taken into consideration. Interestingly, Fellegy (1995) comes to the same conclusion, and Reid (1995) also acknowledges the role of the interlocutor as important, but neither integrate this in their quantifications. Murphy (2012) posits that considering speaker roles, relationships, and other contextual factors is only possible through a careful qualitative analysis of the data.

Murphy's analysis is an excellent demonstration why we need a metric that is sensitive to interlocutor behaviour. In her study, the gender differences are explained through differences in speaker behaviour: in the male group, there are a few men who mostly talk, and a few men who mostly listen. If response token frequency for each speaker is calculated as number of responses the speaker produces per million words in the corpus this leads to extremely high frequency values for those who mostly listen, and extremely low frequency values for those who mostly speak. What these numbers tell us is which interactional role a person has, or rather how balanced the roles are between the two speakers. Consequently, we need a basis for quantification that does not rely on all words spoken by everybody in the corpus, but rather on the behaviour of a speaker relative to their interlocutor.

In the 1970s, social psychologists Duncan and Niederehe (1974) and Duncan and Fiske (1977) presented backchannel frequency normalised relative to 100 interlocutor words in the interaction. Even though this is based on overall counts across a full conversation, and not on a turn-by-turn perspective, it comes much closer to the sort of metric Murphy (2012) as well as Reid (1995) and Fellegy (1995) are implicitly suggesting as a solution to the challenges they have encountered.

To summarise, quantitative work on the overall frequency of LISTENER RE-SPONSES has to meet two main challenges: firstly, the varying definitions and delimitations of the variable mean that the studies are all looking at slightly different subsets of the phenomenon. Secondly, there are many different approaches

to operationalising frequency, but none that takes into account the interactional context in which the variable is produced, leading to potentially skewed results. Thus, we need (1) a definition of LISTENER RESPONSES based on the sequential, interactional context they stand in, and (2) a way of operationalising frequency that takes this context into account.

2.2.2 Qualitative work on Listener Responses

Conversation Analysts have voiced their critique of the first analyses of backchannels early on for treating such a diverse set of actions as 'the same'. Schegloff (1982) pointedly summarises this concern and implicitly suggests a way of addressing it: 'The treatment of them [Listener Responses] in the aggregate, separated from the talk immediately preceding them, loses what they are doing.' (see also discussion around Drummond and Hopper (1993a)). Sorjonen (2001: 285) is even more explicit: 'The type of action to which a response particle is responding is to be treated as part of its interpretation.'

This means we need to draw on the next-turn proof-procedure and the participants' own orientations in the interaction when developing definitions of the functional variants of LISTENER RESPONSES. We can do so by attending to the talk that precedes and follows the individual response, and include their relevant characteristics in our descriptions of the variants, as visually summarised in figure 2.1.

Makes set of potential next actions **relevant**

Preceding Increment

Listener Response

Following Increment

Displays **orientation to / understanding of** prior utterance
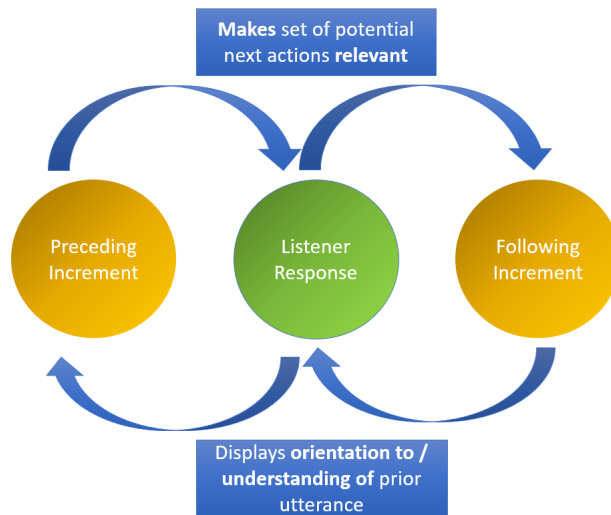
FIGURE 2.1
VISUAL REPRESENTATION OF THE PROPOSED STRUCTURE-BASED DEFINITION OF LISTENER RESPONSES.

This speaks to the need to develop clear, structurally rooted definitions of the variants (see Labov (1963), Schegloff (1993), Heritage (1999), and Pichler (2010)), which can only be done based on a close qualitative analysis and understanding of the interactional processes at hand. However, there currently is no holistic overview of all the actions Listeners can do – possibly because CA does not aim to quantify in this way and it has thus simply not been relevant to develop one. Here, I will draw on three different strands of literature to suggest a basis for a taxonomy and coding scheme.

The three strands of work are: (1) work on individual response tokens or small subsets thereof, focusing on the numerous different functions they can do depending on sequential context and realisation (Schegloff 1982; Jefferson 1984a; Goodwin 1986b; Goodwin and Goodwin 1987b; Barth-Weingarten 2011; Golato and Fagyal 2008). (2) Work on Listenership more broadly, focussing on different actions Listeners do in different contexts and looking at a broader range of response tokens (Goodwin 1984; Gardner 2001; Sorjonen 2001). And (3) work on individual actions that sometimes can be used by Listeners in their responses, for example *assessments* (Pomerantz 1984; Goodwin and Goodwin 1987b), *repair* (see mentions in Gardner (2001) and Schegloff (1982)), or *collaborative completions* (also mentioned in Gardner (2001), referring back to Lerner (2004b) and Lerner (2004a)).

Importantly, I define LISTENER RESPONSES on the level of turn-taking organisation as the things a Listener can do in the Speaker's interactional space without ending the ongoing multi-unit-turn. Thus, they signal continued listenership (or willingness to continue listening, if you wish) in some way. They

- occur between two TCUs of a multi-unit turn or in overlap with a TCU,

- do *not* interrupt the flow of the ongoing TCU, and

- do *not* initiate a speaker change,

This list of criteria is built on the descriptions of *continuers* – which exclusively signal Listenership and invite the Speaker to keep speaking – and *acknowledgements*, which can also be placed turn-initially and signal receipt of the previous talk, while simultaneously serving the Listener to take the floor (Schegloff 1982; Jefferson 1984a; Jefferson 1993).[8] LISTENER RESPONSES are thus defined based on *sequential criteria*, not on form. However, we can note that they are (mostly)

---

[8]Particularly *yeah* and *mhm* have been analysed with respect to the many functions they can serve (Jefferson 1984a; Jefferson 1993), notably the difference between indicating listenership vs speakership incipiency (see the extensive discussion in the special issue of ROLSI from 1993 around the status of *yeah* (Drummond and Hopper 1993a; Drummond and Hopper 1993b; Zimmerman 1993; Tracy 1993; Wieder 1993)).

brief lexical or non-lexical vocalisations or embodied actions, and (usually) have little semantic content. The work looking at individual tokens and their manifold interactional functions takes this observation as a starting point.

2.2.2.1 Qualitative work on individual tokens

Most Listener Responses are brief, often only consisting of a single word or vocalisation like *mhm, yeah, uh-huh, okay, right* or – to give some German examples – *achja, jaja, ja* (Barth-Weingarten 2011; Betz and Golato 2008). These vocalisations can also serve functions other than Listening, and even within the structural scope of a Listener Response, they can serve different functions. For those lexically ambiguous tokens, prosody and its co-occurrence with other cues distinguish which action is being done. I will return to this point in chapter 7.

Listeners can signal *acknowledgement* or *initiate repair* by requesting more information (Schegloff 1982; Gardner 2001), and indicate a *change of state* (produce a 'surprise mark'). All of these immediately hand back the floor to the Speaker (Schegloff 1982; Goodwin 1986b; Stivers 2008; Norrick 2012; Gardner 2001). In his original critique of quantitative work on backchannels that does not distinguish between different functions, Schegloff (1982) differentiates between *continuers* and *acknowledgements*, with *continuers* being the ones that exclusively trigger further talk by the main Speaker, while *acknowledgements* can also signal speakership-incipiency and lead to a floor change. Schegloff (1982) also notes that *repair initiation* is one of the few actions that are always potentially relevant – the Listener can orient to almost anything in the ongoing talk as repairable for some reason, which can then lead to a request for information. This, too, triggers more talk from the main speaker.

Interestingly, Schegloff (1982) subsumes sentence completions, requests for clarification, and brief restatements under the label of 'continuers'; work building on his initial critique introduces further differentiations. Gardner (2001) and Norrick (2012) note *requests for information* as a separate action type that can be done with brief questions or by marking a minimal token with rising intonation. Gardner (2001) further notes *mhm* and *uh-huh* as the most frequent prototypical continuer-acknowledgements, and *right, yeah* and *okay* as the most prototypical acknowledgements (ie those that can also signal speakership incipiency).

Schegloff (1982: 85) also gestures at *markers of surprise*, often called a 'change-of-state' token, as a possible Listener Response – here the Listener indicates that they now know something they did not know before. This type of response often leads to the Speaker elaborating more on what they have just said (Tolins and Fox Tree 2014). The most prototypical and most discussed surprise mark is probably

Oh (Heritage 1984; Schiffrin 1987a), but exclamations like *Really?* or confirmatory tokens like *Right* can do the same job (Gardner 2001).

2.2.2.2 Listenership more broadly, and actions that can sometimes by Listener Responses

Scholars focusing on Listenership and specific actions Listeners can do point out that in addition to the actions listed above, Listeners can also do *assessments* (particularly Goodwin (1986a) and Goodwin and Goodwin (1987b) and the monographs by Sorjonen (2001) and Gardner (2001)). Gardner (2001) further notes, with reference to Lerner (2004b) and Lerner (2004a), that they can *co-construct* the Speaker's turn without taking over the floor.

Assessments are contributions that in some way evaluate (an aspect of) the preceding talk. *First assessments* orient to something that was said as an assessable, while second assessments orient to a preceding assessment and (dis)agree with it (Pomerantz 1984). Schegloff (1982), Goodwin (1986a), Goodwin (1986b), Goodwin and Goodwin (1987b), Jefferson (1993), Norrick (2012), and Gardner (2001) have described first assessments as possible Listener Responses, both minimal and more elaborated ones following the format described by Pomerantz (1984).

Interestingly, *second assessments* do not seem to have been investigated yet. Structurally, they should be possible as Listener Response – if the Speaker makes a first assessment, the Listener responds to this, and then the Speaker continues. Their occurrence might be rare, however, given the (at least potential) closing-implicativeness of assessments (see Antaki et al. (2000), Goodwin and Goodwin (1987b), Lindström and Heinemann (2009), and Mondada (2009)).

Additionally, Gardner (2001) mentions joint utterances as possible actions Listeners can do, building on Lerner (2004b) and Lerner (2004a). Particularly in involved narratives, Listeners can enter the main Speaker's turn-space and contribute to an ongoing turn, without actually taking over the floor. These collaborative contributions are often marked by lower pitch, a slightly breathy voice quality, and their brevity. Most of these contributions are not 'necessary', they are joint utterances that could be taken to express alignment and affiliation (Stivers 2008). However, it is also imaginable that the Speaker indicates word-search trouble in some way, thus initiating repair, and the Listener supplies the word in question. This makes *self-initiated other-repair* another possible Listener response action, and a potential sub-type of joint utterances.

2.2.2.3 Summary of the Action Types

To summarise, this review provides a cohesive overview of the different actions Listeners have been described as doing. I suggest defining the variable *structurally and sequentially* as all those things Listeners can do without taking the floor. Such a structured overview also highlights potential structural gaps. The action types derived from the literature reviewed are: (1) acknowledgements, (2) surprise marks, (3) assessments, from which so far only first assessments have been analysed; I suggest (4) second assessments can be Listener Responses too. Further, Listeners can (5) initiate repair through a request for clarification or elaboration, and (6) co-construct the Speaker's turn. Here, only collaborative completions that are not directly elicited by the Speaker have been attested. I would like to suggest that structurally it is also possible to have (7) self-initiated other-repair as a Listener Response, i.e. word-search completions. This review also reminds us of the importance of considering prosody as part of the form, as mentioned in the theoretical section on non-quantitative CA work on variation.

Chapter 5 tests and further develops this taxonomy based on just over 5200 single cases analysed in this thesis, and chapters 6 and 7 show the potential of applying such a structurally-based definition of the variable and the variants to quantitative analyses.

2.2.3 Mixed Methods work on Listener Responses

As we have seen in the theoretical section on defining variables above the level of the phoneme and their variants, there are several key problems if we do not take an interactional approach: (1) Researcher-imposed top-down definitions vary between studies and researchers, (2) sometimes structural constraints are included, but if and how this is done varies, and (3) there are next to no inferential models for within-category variation, especially none taking into account structural constraints.

We can see this reflected in the mixed methods work on LISTENER RESPONSES: the sub-categories or functional variants described vary between different studies, and even though some studies try to take into account structural constraints, there is no cohesive approach with respect to structure either. Some studies further distinguish form-based variants, classifying LISTENER RESPONSES based on their length or morphological complexity.

2.2.3.1 Variation in function-based coding

Function-based variants range from a binary distinction to five sub-categories. Bavelas et al. (2000) distinguish between *generic* and *specific* responses, where generic ones only orient to the fact that the other person is talking and signal listenership in the broadest sense, while *specific listener responses* display an orientation to *what* the speaker is saying and include exclamations, assessments, or other reactions that evaluate the preceding talk in some way. This distinction has proven productive and has been picked up by other researchers looking at listener responses (Guardiola et al. 2012; Guardiola and Bertrand 2013; Tolins and Fox Tree 2014).

Reid (1995) also makes a two-way distinction, but she draws the line between *facilitative* and *agreement* backchannels. These are different categories from the ones introduced by Bavelas and colleagues – at first sight *agreement* might sound similar to *specific* listener responses, but the *specific responses* include all kinds of stances, not only agreement, and *facilitative responses* can also display a clear orientation to the preceding talk. To my knowledge, Reid's distinction has not found much further application.

Other scholars more closely rely on interactional work when defining functional categories: Kogure (2003) bases his five-way distinction on Clancy et al. (1996), coding for *continuers, reactive expressions, repetitions, collaborative finishes*, and *resumptive openers*. These last ones lead to a change in speakership and are thus not things Listeners can do and remain Listeners. O'Keeffe and Adolphs (2008) make a four-way distinction between *continuer, convergence, engagement*, and *information receipt tokens*, following Schegloff (1982).

Some studies find variation in the frequency of different functions by gender. Stubbe (1998) for example compares verbal LISTENER RESPONSES in English-language interactions between male and female New Zealanders with Pakeha as compared to Maori roots. She finds that overall men produce more verbal feedback than women. Once the responses are categorised into *neutral minimal response*, *supportive minimal response*, and *(supportive) cooperative overlap*, women are found to produce more *supportive* verbal feedback, while for men a higher percentage of their LISTENER RESPONSES is considered *neutral*. However, she does not take into account opportunities to vary: the respective interlocutors might make different types of LISTENER RESPONSES relevant.

2.2.3.2 Variation in form-based coding

Some scholars also assume a form-function link and therefore break down the data based on form. For example, Guardiola et al. (2012), who draw on Bavelas

et al.'s (2000) distinction between *generic* and *specific* responses, apply a measure of 'morphosyntactic richness' based on word classes. They argue that *specific* backchannels are morphosyntactically richer than *generic* ones and find that morphosyntactic richness increases over the course of a narrative in their data.

McCarthy (2003) determines a set of LISTENER RESPONSES on a quantitative basis: He draws on 3.5 million-word subsets of one American and one British corpus of spoken interaction to create a list of 19 tokens (McCarthy 2003: 47) that occur at least 100 times in each. He then categorises the tokens as part of one of six environments with respect to form. The most frequent one is that of a single-word, lone-standing LISTENER RESPONSE, but they can also precede expanded responses or be followed by a modifier. Further, the tokens can be negated, occur in doublets / triplets, or longer chains, or they can stand in clusters or extended sequences. Interestingly, McCarthy (2003) does not offer any frequency distribution for these six categories, and it is not clear which analytic method he uses to derive them from the data. These categories cut across different actions Listeners might do and are solely form-based.

Two studies find no difference in how often male or female interactants produce LISTENER RESPONSES, but variation on the level of form of the response: Oreström (1983), working with a subset of the London-Lund-Corpus, finds men and women produce similar numbers of LISTENER RESPONSES but use different tokens and vary in their prosodic design. Women use *mm hm* and *yeah* less frequently than men, and the token *yes* tends to be produced with falling intonation by women, but with a flat pitch contour by men. Similarly, Maltz and Borker (1983) find that men and women use the same token to do different actions, and Kjellmer (2009) notes that men and women produce LISTENER RESPONSES at the same overall frequency in the Cobuild Corpus, but that they draw on different lexical material.

2.2.3.3 Variation in placement-based coding

There is huge variation with respect to placement-based coding in those studies that do consider it. In conjunction they show that there are places at which LISTENER RESPONSES *can* relevantly be produced, and that certain contextualisation cues make them especially relevant, but at the same time they do not *have to be* produced at every point of possible occurrence. Even though I do not analyse LISTENER RESPONSES with respect to their placement in this thesis, I present a brief overview of studies that do for completeness sake, and to note that this might be an interesting avenue for future work.

Some scholars annotate whether the response occurs in overlap with the ongoing talk or not (Heinz 2003), while others distinguish between LISTENER RE-

SPONSES at a TRP, at the end of a TCU, elsewhere, or in the structural position of an answer (Fellegy 1995; Reid 1995; Kjellmer 2009). Others define up to 12 different 'opportunities' to produce a backchannel (White 1989: 64), related amongst others to the pitch of the preceding talk, overlap, triggering particles (or other triggering behaviours like gaze), Turn Constructional Units (TCUs) or Transition Relevance Places (TRPs).

Computational work tries to predict the occurrence of backchannels based on pitch, with regions of low pitch often preceding listener responses (Levitan et al. 2015; Heldner et al. 2010; Cathcart et al. 2003; Ward 2006), or pauses (Fellegy 1995; Levin and Lin 1988; Cathcart et al. 2003) in different varieties of English as well as in Japanese. Other scholars show that small particles (Maynard 1990) often trigger a listener response, or that gaze can have the same effect (Bavelas et al. 2002).

Some scholars find variation by gender in terms of placement: In the data presented by Fellegy (1995), 95 % of all backchannels occur at phrase boundaries for both genders, but women spread their responses across different places of occurrence (within turns, at the end of a multi-unit-turn, or elsewhere) while men backchannel mostly within turns. What goes largely uncommented in this study is that women produce three times as many LISTENER RESPONSES as men in the same amount of time.

Guthrie (1997) combines considerations of function, form, and structure in her study, without looking at social factors: Focusing on two specific tokens, *okay* and *mmhmm*, she illustrates differences in their distribution and function using Conversation Analysis in concert with some basic frequency counts. Guthrie observes that *okay* and *mmhmm* both often stand in 3rd-turn position as receipts of previous talk and seem interchangeable. An initial sequential analysis shows that the two tokens occur in three different positions: (1) Lerner-type compound TCUs, (2) places of possible syntactic completion only, or (3) places of possible syntactic and prosodic completion. However, they occur in these positions at different frequencies. Guthrie (1997) then presents a 'deviant case' analysis of extracts that do not conform with the result of this quantification, showing there is still more to be explored. Even though the quantification (raw frequency counts and a likelihood test) is not overly complex, this is an example of how qualitative and quantitative analysis can work in concert.

One study that successfully intertwines sequential and basic inferential statistical analysis with respect to placement and co-occurrence patterns is Hömke et al. (2017). They clearly situate the variable in its sequential and multimodal context and show that eye blinking can be listener feedback. With respect to structure, Hömke et al. (2017) annotate their data for final and non-final TCUs, following Ford and Thompson's (1996) description of complex TRPs and Ford and Thompson's

(1996) and Lerner's (1996) notion of local pragmatic completion. Multimodal cues like nods, vocal continuers, combinations of both, and gaze (mutual vs. not mutual) were also annotated. The eye blinks were then split into long and short ones based on quartiles (the shorter 75% have a duration of less than 410ms, and long ones in the upper quartile are longer than 410ms).

In terms of statistical analysis, Hömke et al. (2017) present a Chi-Squared test that shows long blinks are more likely to co-occur with head nods or continuers than short blinks. Based on this, the authors randomly select 61 short blinks to compare to the 61 long blinks and find in a combination of quantitative and interactional analysis that long blinks are restricted to specific sequential locations and occur there far more frequently than short ones. The authors conclude with a qualitative, sequential analysis of two extracts that illustrate how eye blinking functions as listener feedback.

This study is an excellent example of how qualitative and quantitative analyses can cross-fertilise and drive each other (similar to the first steps taken in Guthrie (1997), but far more developed in terms of the statistical analysis), and a potential blueprint for further mixed methods studies involving CA.

2.2.3.4 A note on gender

I have focussed on studies that take gender as a social variable where possible. I follow these studies in taking a strategic essentialist view of gender by treating it as a binary social variable (see the excellent discussions in Meyerhoff and Stanford (2015) and Meyerhoff and Ehrlich (2019)). Of course this flattens reality – just as applying a coding scheme to interaction does – but it allows us to build on previous work, relate to broader discussions in work on language variation and gender, and drive forward an exciting theoretical and methodological project.

I align with the interaction-focussed critiques of the 'dominance' and 'difference' approaches to language and gender that have been formulated with respect to other discourse-pragmatic (but also phonological) variables (Eckert and McConnell-Ginet 2003c; Meyerhoff and Ehrlich 2019). The 'dominance' approach posits that gender differences in language reflect social power differences between the sexes. It assumes men and women use different linguistic codes ('men's/women's language') which both reflect and reinforce women's position as the 'weaker' sex (see Lakoff (1973)). Numerous studies have shown that 'power' is performed and negotiated locally in interactions, and that there are far more differences *within* each group than *between* men and women.

The 'difference' approach equates sex with gender, and posits that perceived systematic physiological differences between men and women lead to different

characteristics and behaviours. This implies that these differences are innate. By now, this perspective has been deconstructed (and debunked) in a number of ways (though some scholars and public discourse still subscribe to such a view): physiological differences between men and women are not as great and clear-cut as often presented, and socialisation is an extremely important factor in any person's development (for an excellent overview see Eckert and McConnell-Ginet (2003a)).

Scholars have found with respect to tag questions, overlap, high-rise terminals, hedges, compliments, turn-taking or even how much any one person talks in an interaction that these actions are contingent not only on the party who produces them, but also on the person they are interacting with, and the interactional projects both embark on together (see the reviews in Holmes (1995), as well as Dubois and Crouch (1975), Holmes (1984), Cameron et al. (1989), and Wilkinson and Kitzinger (2014) and the work by Eckert and McConnell-Ginet as well as Meyerhoff and Ehrlich cited above).

For all the variables above, interactional structure and both parties' behaviours are relevant in (1) providing a definition of the variable and the envelope of varia-tion, (2) the variants, and (3) fully understanding how the variation observed comes about, and what it interactionally means. Meyerhoff (2014: 100) eloquently argues for the necessity of integrating social dialectology and discourse analysis if the goal is to understand the meaning of any specific form or realisation at the very local level of the interaction, and then to understand how this relates to broader social patterns.

Interestingly enough, this awareness has not been applied to analyses of backchannels or LISTENER RESPONSES in depth, as Eckert and McConnell-Ginet (2003b) note: they cite three studies that find women producing more backchan-nels than men (Bilous and Krauss 1988; Roger and Nesshoever 1987; Edelsky and Adams 1990), and contrast this with Maltz and Borker (1983) who argues that in fact men and women use individual response tokens to do different actions (signalling attention vs agreement). This ties in with the broader CA critique of work on LISTENER RESPONSE frequency as conflating the complexity of such a diverse set of tokens and actions.

So, while this thesis will primarily focus on the theoretical contributions out-lined next, it also contributes to the body of work on socially situated analyses of language variation and gender in two ways: first of all, by applying knowledge gained from the analysis of other discourse-pragmatic variables to LISTENER RE-SPONSES, and secondly by showing that we need an even greater awareness of *how* exactly interaction and individual actions are co-constructed before we ascribe them to any one party.

2.3   Joining the dots – contributions of this thesis

Currently, sociolinguistic theory and methodology does not include interactional structure into defining the variables, variants, and the approach to quantification. We have seen that this is problematic, both on an abstract-theoretical level in the first half of the literature review, and exemplified it practically with respect to LISTENER RESPONSES in the second half. Specifically, the complications relate to (1) working with variables above the level of the phoneme, and (2) understanding the relationship between phonetic and prosodic variation with interactional structure.

To address these challenges, we need to develop (1) an interactionally-rooted definition of the variable, the envelope of variation, and an accountable way of quantifying, (2) a coding scheme for the variants that situates them in the interactional structure and thus allows for an analysis of structural constraints, and (3) a way of quantitatively analysing the variation observed based on both social and structural variables that goes beyond frequency distributions. With respect to the second bigger question, we need (4) an approach that allows us to connect the level of the interaction – i.e. action types or functional categories – to the level of the linguistic and multimodal realisation of each utterance.

I address these gaps in the four analysis chapters – the first takes up the issue of quantification, the second presents a taxonomy of LISTENER RESPONSE actions, the third integrates this coding into the quantitative analysis, and the fourth tackles the question of form and function. In all chapters, the notion of sequential structure and the next-turn proof-procedure are crucial to the analysis. I suggest quantifying overall frequency on a turn-by-turn level, relative to the amount of speaker-talk that is being responded to (chapter 4). The description of the individual variants or action types in chapter 5 is also rooted in interactional structure, building on the taxonomy drawn from the literature that I have presented here. In chapters 6 and 7 I show that the sequential context and surrounding talk are important predictors for variation and introduce inferential statistics for analysing variation between the action types.

Thus, in the next chapters I will demonstrate based on the example of LISTENER RESPONSES and GENDER that integrating Conversation Analytic tools and methods to the quantitative analysis of variation can resolve ongoing theoretical and methodological quandaries with respect to variables above the level of the phoneme.

Chapter 3

Methodology

> Man tar lite härifrån och lite därifrån och joxar ihop det så blir det lagom jox.
> –
> *You take a bit from here, and a bit from there, and mix it, so that in the end it's just the right blend.*
>
> *Astrid Lindgren*

The first part of this chapter presents the data that form the basis of this thesis, including reflections on the implications of the study design, the second briefly summarises the qualitative definition of Listener Responses and the tools used in the analysis, and the third and final section shows that inter coder reliability was exceptionally high for all levels of coding, demonstrating that the integration of interactional and quantitative methods can result in reliable, highly reproducible coding schemes.

3.1 The Data

I first outline the study design and participant recruitment, including ethical considerations, then discuss data collection and data management, give an overview of the selection of data I draw on in my analysis, and end with a discussion of some practical implications of the study design.

3.1.1 Study Design

This PhD dissertation is part of a bigger study on the experience of people living with (or caring for someone living with) Type 1 Diabetes in Scotland. The goal of the wider project is to collect best practices and challenges in a number of

areas pertaining to Type 1 Diabetes management, including support provided by the NHS and healthcare professionals. Within this bigger project, three types of data were collected: a set of questionnaires including a sociolinguistic background questionnaire (Appendix C), three rounds of dyadic interactions, and one interview with each participant.[1] I will describe the setup of the dyadic interactions here, and outline the overall structure of each data collection session in a separate subsection.

Each participant took part in three topic-focused dyadic interactions. For each person, the the first topic was 'Diabetes technology and healthcare', the second one 'Diabetes in your daily life', and the third 'Diabetes and mental health'. At the start of each conversation, the respective topic was introduced as the broad focus of the interaction, and participants were invited to discuss anything relating to it. To facilitate the interactions, I had prepared a set of prompts for each round and offered these to the participants as inspiration in case they were unsure where to start. I explained that using these prompts was optional. Nevertheless, most dyads referred to several or all of them. The prompts consist primarily of cartoons, comics, or pictures of items or situations related to Type 1 Diabetes, as well as some brief verbal prompts. The full set is attached in appendix D.

3.1.2   Data Collection

In this section, I describe the data collection process for each round of recordings, summarised in table 3.1. Each data collection session includes four participants involved in different activities at different times. It is split into a pre-conversation stage, a conversation stage, and a post-conversation stage, with participants arriving at two different times to minimise the time commitment for each person. Each of the blocks listed above takes approximately 30 minutes. With 5 blocks each person takes part in, this adds up to a time commitment of 150 minutes. Participants were compensated with GBP 20 for their time.

Stereo-recordings with a separate track for each speaker were made using a Marantz PMD661 recorder (for the first two sets of recordings) or a Zoom H5 recorder (for the second two sets of recordings). Speakers wore head-mounted

---

[1] Given the specific set of participants and the content of their conversations living with Type 1 Diabetes and experiences with (access to) health care here in Scotland the additional questionnaires were a Perceived Stress Scale (Cohen et al. 1983), the Oxford Happiness Questionnaire (Hills and Argyle 2002), a short questionnaire checking for the big 5 personality factors (Saucier 1994), as well as the self-consciousness scale in the revised version for non-student populations (Scheier and Carver 1985). All these pieces of information might be relevant in the context of the broader project on how people talk about their experiences of living with Type 1 Diabetes, but they are not crucial to the linguistic analysis presented here.

microphones (Shure SM10a). Recordings lasted from 16 to 45 Minutes, adding up to 17 hours of data in total.

Round	Room 1	Room 2	Topic
Start: A and B arrive, sign consent form (appendix A)			
Pre 1	Interview: A	Questionnaire: B	Interview or Questionnaire
Pre 2	Interview: B	Questionnaire: A	Interview or Questionnaire
C and D arrive			
Round 1	A+B	C+D	Diabetes technology and healthcare
Round 2	A+C	B+D	Diabetes in your daily life
Round 3	A+D	B+C	Diabetes and mental health
A and B are debriefed, receive their compensation, and leave			
Post 1	Interview: C	Questionnaire: D	Interview or Questionnaire
Post 2	Interview: D	Questionnaire: C	Interview of Questionnaire
C and D are debriefed (appendix B), receive their compensation, and leave			

TABLE 3.1
SCHEMATIC REPRESENTATION OF ONE DATA COLLECTION SESSION, WITH
PARTICIPANTS REFERRED TO AS A, B, C, AND D

In the schematic representation of one data collection session (table 3.1) the four participants are referred to with letters. The complexity of this so called round robin design becomes evident: data collection happens in two rooms simultaneously, and participants arrive at different times. Upon arrival, the first two participants are welcomed, informed about the general purpose of the study, and given consent forms. The data collection process was approved by the *Linguistics and English Language Ethics committee*. In terms of anonymising the data, the participants were offered to decide on their own pseudonym, ask for their real name to be used, or ask the researcher to invent a pseudonym. The consent form is given in Appendix B.

The first two participants to arrive take turns being interviewed and filling in the background questionnaires during the first two pre-sessions. Once the other two participants have arrived and signed the consent forms, the core recording round starts. Participants rotate rooms, so that in the three rounds each person talks to each other person in their group of four. For the two participants who arrived early, the data collection is complete after the last topic-focused conversation; they are debriefed, informed of the linguistic research questions pertaining to the project, and asked to renew their consent (see Appendix F). The two participants who arrived just before the topic-focused conversations now take turns giving an interview and filling in the background questionnaire. They are then

also debriefed and asked to renew their consent at the end of their stretch of the recording session.

Four recording sessions took place; three in the linguistics department at the University of Edinburgh, and one in the linguistics department at the University of Glasgow. The rooms used in Edinburgh are small meeting rooms with a table and seating for up to six people. In order to create a more relaxed atmosphere, tea, coffee, and snacks were arranged on the table, as were the colourful prompts for the conversations. One of the meeting rooms contains a small philosophy library, which makes this room a bit more like a living room, while the other one is more neutral. Because there was construction work around the building at the time of recordings, some of the noise can be heard in the background of the conversations. Recordings in Glasgow were made using the Glasgow University Laboratory of Phonetics sound booth or a professor's office. Again, the sound booth is more neutral, while the office was much more personalised. Tea and snacks were arranged on the tables here, too, but participants did not usually eat during the recordings, rendering the audio tracks mostly clear and without interference.

I will briefly describe the interview process, because I draw on the information obtained during the interviews, even though I do not analyse them with respect to LISTENER RESPONSES. The interviewer for all participants was Daisy Smith, a fellow PhD student from the Newcastle area whose research focuses on Older Scots. She is hence familiar with Northern and Scottish varieties of English and a knowledgeable outsider when it comes to Type 1 Diabetes. The semi-structured interviews were based on a set of questions focusing on the experience of being diagnosed, the support participants received, and the change and impact they hope for. The questions are attached in Appendix E. I chose an outsider as the interviewer and added the interview format, because this social situation leads to different kinds of Speakership and Listenership: not as much shared knowledge or common ground can be assumed with someone who does not live with Type 1 Diabetes, which I hypothesised would lead the participants to explain and elaborate more than if I had interviewed them myself. Given the scope of this PhD thesis, the interviews were not analysed with respect to Daisy's and the interviewee's behaviours as Listener and Speaker, but it would be a very interesting follow-up study to directly compare the same participants in two different interactional situations.

3.1.2.1 Recruitment

I recruited participants through Diabetes Scotland, Facebook groups related to Type 1 Diabetes in Scotland, and the Twitter Diabetes online community. The advertisement I shared was titled *Talk about Type One: the Highs and the Lows* (Appendix A) and contained some basic information on the study, framing it as research into living with Type 1 Diabetes in Scotland. Over 40 people registered their interest within only a few weeks. Because of the complex scheduling requirements (4 participants needing to come to the same place at the same time for approximately three hours), I arranged recording sessions based on a doodle poll and chose those participants that naturally formed a group of four based on their availability. It is worth noting that I had no problems at all with cancellations or no-shows. I would argue this is because the participants genuinely cared and had an intrinsic interest in discussing the topic and contributing to the conversations and the larger study on living with Type 1 Diabetes in Scotland.

3.1.2.2 Ethical considerations

Data collection for this study was approved by the University of Edinburgh PPLS Ethics Committee on 22 November 2016. As mentioned above, participants were recruited under the headline of 'Talk about Type 1', while the focus of this PhD thesis is on LISTENER RESPONSES. It is worth noting that not a single participant suspected that part of the analysis would focus on *how* they were talking rather than just *what* they were saying. This could be considered deception and therefore problematic. However, the PhD thesis is part of a larger project interested in how people living with or caring for people living with Type 1 Diabetes talk about their condition, and which topics are important to them. Furthermore, the participants all said they enjoyed and benefited from the conversations for their own sake, and would have participated even if they had not been financially compensated for their time.

For this broader project, I collected additional materials not used in this thesis (outlined above), as well as a Twitter corpus comparable in size to the conversations analysed here. I used this Twitter corpus to assess whether the topics raised by the participants in my study were representative of the discussions we can observe in the broader Type 1 Diabetes community.[2]

---

[2]So far, posters, talks, and drafts of papers have come out of the broader project, I will focus on publishing results of the qualitative and quantitative content analysis and disseminating these findings in the community after submission of the PhD.

3.1.3 Data Management

3.1.3.1 Filing and Naming system

There is one audio file per conversation with a separate recording channel for each participant, one time-aligned ELAN transcription file that also contains the annotations and coding of all Listener Responses, and one associated praat text grid for each conversation. Filenames are structured consistently as *Date_Name1-Name2_Topic.** to allow automatic matching during the processing phase.

3.1.3.2 Transcription in ELAN (and Praat)

ELAN versions 5.1 and higher (Sloetjes and Wittenburg 2008) were used for transcription and annotation, following the FAVE transcription guidelines (Rosenfelder et al. 2014), with one tier per speaker. Listener Responses were annotated on a separate tier for each speaker, and coded individually for their function on a third tier. I had to allow for overlapping annotations in order to keep the time-alignment accurate for each participant, leading to the tier structure represented in table 3.2.[3] 'Stereotypes' in ELAN define the relationship between two tiers that are dependent on each other. The stereotype *Included In* is described as follows in the ELAN manual (Hellwig et al. 2018: 195): 'All annotations fall within the borders of the parent tier. However, there can be gaps between the child annotations. E.g., a sentence with a silence can be split up into words while the silence corresponds to a gap in the child annotations (i.e. the separate words).' The stereotype *Symbolic Association* is described as follows: 'The annotation on the parent tier cannot be sub-divided further, i.e., there is a one-to-one correspondence between the parent annotation and its referring annotation.'

For qualitative analysis, individual similar case collections were created, one for each action type category. CA transcripts were created in Praat versions 6.0.36 and higher (Boersma and Weenink 2016) and exported to .txt for further editing.

3.1.3.3 Processing and Analysis

All conversations were coded in ELAN based on the qualitative coding scheme described in chapter 5. 10% of the data were coded by a second coder, and inter-coder-reliability was calculated. The coding manual and workflow used are given in appendix H. Data from all 24 annotated conversations were exported from ELAN

---

[3]Note that in terms of the amount and complexity of processing required after exporting the ELAN files to CSVs and minimising the number of tiers required, it would have been far easier to create only one speech tier, one 'Speaker' tier on which the speech would be assigned to the person talking, one 'Listener Response' tier, and one 'Action Type' tier. I decided against this, in order to create as accurately a representation of the actual development of each interaction as possible.

Level 1	Level 2	Level 3	Tier content	Stereotype
S1 Talk			verbatim transcript for Speaker 1	included in
	S1 Listener Responses		Listener Responses by Speaker 1	included in
		code	Action Type	symbolic association
S2 Talk			verbatim transcript for Speaker 2	included in
	S2 Listener Responses		Listener Responses by Speaker 2	included in
		code	Action Type	symbolic association

TABLE 3.2
ELAN TIER STRUCTURE

in a CSV file that can be read into R (R Core Team 2016). All processing and analysis was done in RStudio (version 1.1.414 and higher, RStudio (2015)), relying primarily on the TIDYVERSE family of packages.

3.1.4 Describing the data

Unlike most sociolinguistic studies, I decided not to control for age, gender, or sociolinguistic background in the recruitment process. Instead, I focused on the common denominator that provided the conversation material: living with Type 1 Diabetes. Given the complex set-up and my need to rely on the participants to actually attend the scheduled recording sessions (if one person had not attended, the round robin design would not have worked), I decided to prioritize engagement with topic over social stratification. Furthermore, living with Type 1 Diabetes as a common denominator provided so much common ground, topics, and empathy that the vast majority of the conversations are very free, natural, and narrative in character, making them a very rich base for the analysis.

3.1.4.1 Social factors

As described above, the sample is a convenience sample. However, it is diverse in terms of region, age, social status, and gender. This allows me to argue that the Listening practices are not restricted to one social group, as well as to look into inter-speaker and inter-group differences. Overall the sample is skewed towards

local Scottish, well-educated, relatively young people, but there is representation beyond young white middle-class speakers. This is a positive surprise given that no social factors were controlled for at recruitment. I will briefly summarise the sample with respect to the social variables of age and gender, and then talk about region and education.

The 16 speakers can be grouped by age and gender as summarised in table 3.3 below. There are more female than male participants, and the sample is skewed towards the younger population (ages range from 19 to 71, mean age is 36). 14 are people living with diabetes, 2 are mothers of children with Type 1 Diabetes.

Age group	Male	Female	Total
19-29	2	4	6
30-39	2	3	5
40-49	0	2	2
50-59	1	1	2
60+	0	1	1
Total	5	11	16

TABLE 3.3
PARTICIPANTS SPLIT BY AGE AND GENDER

None of my participants self-identifies as non-binary with respect to their gender, and gender is not foregrounded in the conversations. Biological sex does come up because of practicalities relating to diabetes management: particularly the women talk about the interaction between insulin sensitivity and hormones (the younger ones with respect to the menstrual cycle, and the two older women also with respect to the menopause), and those who wear insulin pumps discuss the challenges of 'hiding' or at least storing the device when wearing dresses (an interesting potential intersection of sex and gender). The question of what to do with the insulin pump during moments of physical intimacy is also briefly touched upon. Thus, biological sex does play a role in terms of practical considerations, but gender is not foregrounded or explicitly discussed in the interactions under scrutiny here.

At the time of recording, all participants resided in Scotland, 9 of them in Edinburgh, 3 in Glasgow, and the other 4 in the Central Belt of Scotland. 11 participants were born and grew up in different parts of Scotland, 5 are from the UK or abroad: One participant is originally from Poland and has been in Scotland for 14 years, one is Finnish-American and has been in Edinburgh for 7 years, and one is Irish and has been in Edinburgh for half a year. One is from the South of

England and moved to Scotland for her studies 10 years ago, and one is from the South West of England and moved up to Scotland almost 20 years ago.

With respect to mobility, those participants who were born in Scotland are either not mobile at all or only moved once or twice within Scotland to study or for work. Those participants who came from outside of Scotland are more mobile and have lived in different parts of the UK and/or different countries. 4 participants have no university education, 6 have undergraduate degrees, and 6 have postgraduate degrees. One of the speakers has a cleft lip (from birth), which influences the interactions between him and the other participants in his group.

Considering the representation of different social variables in my data, I chose to analyse variation based on gender. It is best represented compared to the other social variables, easy to operationalise in the context of the recording, and provides enough tokens per group to allow for complex statistical analyses.[4] The methodological and theoretical approaches presented here can be applied to any other social variable; gender is simply used to convey this methodology and because there is a comparatively substantial body of work on gender-based variation in Listener Responses to be discussed.

3.1.4.2 Total talk and tokens

A total of 5202 Listener Responses were identified and coded. Those utterances that were not coded as Listener Responses were tagged as 'talk' in order to count up the number of words in a turn, with 'turn' being operationalised as 'the number of 'talk' words from Speaker change to Speaker change'. Recordings were on average 26 minutes long (sd = 6:20 min), ranging from 15:40 to 40 minutes. Table 3.4 shows the mean, sd, and range for the number of words and Listener Responses per participant per conversation. More detailed breakdowns and summaries by group will be presented in the respective analysis chapters. A more extensive overview of recording durations, word counts and Listener Response counts by conversation and by participant is included in Appendix I.

	Mean	sd	Min	Max
Listener Responses	108	46	26	222
Number of words	2350	882	909	4558

TABLE 3.4
SUMMARY STATISTICS FOR NUMBER OF LISTENER RESPONSES AND
NUMBER OF WORDS UTTERED PER PARTICIPANT PER INTERACTION

---

[4]As pointed out above, I acknowledge that gender is not simple and binary but rather performed and negotiated. However, for the purposes of this study, a binary classification of sex was a practical metric to be used in the analysis.

3.1.5 Implications of the Study Design

The study design and data collection described above have several implications for the generalisability of the findings, as well as for the frequency and types of LISTENER RESPONSES observed. While it is important to acknowledge these implications, the goal and contribution of the present thesis is a theoretical and methodological one. This situates its generalisability in the potential to transfer the integration of interactional and quantitative methods to other variables and analyses rather than at the level of findings about individual categories.

The two core challenges LVC researchers will notice are first that recruiting for 'Talk about Type 1' resulted in an unbalanced sample, with approximately twice as many women than men. Second, there is only one recording context, and in this specific context we can expect that the interactants are 'doing being' good participants.[5] While descriptive and inferential statistics make it possible to detect trends and compare across groups, we still need to be cautious not to overextend the findings to the wider population or other situations. On the other hand, the resulting dyadic interactions, recorded at high quality on separate channels for each Speaker, are well-suited for linguistic and interactional analysis: focussing the participants' attention on a shared topic rather than the recording situation is an innovative approach to overcoming the observer's paradox and results in very spontaneous, naturalistic speech. In contrast, most previous studies of language variation have relied on sociolinguistic interviews (see Pichler (2013b) or comparatively staged conversations, for example between faculty and students (White 1989)).

From an interactional perspective, we need to take into account (1) that the recordings stem from one context only, (2) that the interactions, though naturalistic, were set up by me and therefore did not occur naturally, and (3) that I only analyse vocalised responses even though the participants interacted with each other multimodally. I discuss the first two challenges together, and then briefly address the third point regarding multimodality.

As stated in the literature review relative to the methodology, when investigating a specific phenomenon CA scholars usually work with collections of 'similar cases' drawn from different individuals and interactional contexts. The given recording situation constrains or at least influences the participants' behaviour in several ways: It is likely that they are more attentive to each other's stories and cues than they would be in everyday interactions, where attention is split between practical tasks, general decision-making about day-to-day life, other concerns, and

---

[5]This phrasing reflects the Conversation Analytic orientation of focussing on the process rather than the result – the participants *are* not intrinsically good participants, but they *perform* these roles, they '*do* being' good participants.

whichever interaction they are currently participating in. Situations where one party has the other's undivided attention are relatively rare and bring out more attentive Listening behaviours (see Bavelas et al. (2000), who found that when Listeners were distracted, they produced more 'generic' responses compared to when they were paying full attention). Thus, in my data we might see more 'specific' responses, while in casual everyday-talk where participants are not as attuned to the conversation 'generic' responses would be more frequent. In a similar vein, we do not find Speakers 'policing' Listeners' lack of engagement or problematising the absence of a relevant specific response in the data under scrutiny, while this is something that happens in day-to-day interactions.

Consequently, not all possible ways of Listening might be included in my data and thus analysis, and the distribution of Action Types cannot be generalised to other contexts. Nevertheless, the action structures and sequences observed are valid behaviours and strategies, and most likely not restricted to the situation or the participants at hand: With respect to the number of cases, CA scholars uphold that one single case is enough when it comes to showing that a phenomenon exists. There are two good reasons why it is unlikely that the strategies described here are not restricted to the setting at hand:

First of all, both parties involved in any given interaction recognise and participate in the ways of Listening and Speaking I observe without interaction breaking down at any point. This indicates that they are familiar with these strategies. If these rules were negotiated on the spot, we would expect more repair-initiations and difficulties. Secondly, none of the participants had met any of their interlocutors previously. This means that the action strategies observed cannot be idiosyncratic behaviours developed between close friends, families, or partners. The fact that all participants immediately recognise and participate in these ways of Listening and Speaking indicate that they are commonly used outside of the particular context I recorded them in.

The third concern is the focus on the vocal modality in a multimodal interactional context. Here, I can only acknowledge this and remind the reader that I do not claim to cover all possible Listener Responses on all levels. The decision to focus on the vocal modality only was taken in order to make the presented integration of interactional and quantitative methodologies possible and to develop a clear, basic model which future work can expand on to include more modalities and contexts. Previous work has found vocalised and non-vocalised responses to co-occur approximately 75% of the time (Duncan and Fiske 1977). This suggests that looking at vocalised responses captures about three quarters of the responses done in the interaction, and reminds us that the absence of a vocalised response does not entail the complete absence of any response at all. Thus, all findings

presented here are with reference to the frequency and functions of *vocalised* Listener Responses only.

To summarise, the study design has allowed me to create high-quality recordings of naturalistic, spontaneous talk-in-interaction in a complex round robin design. This comes at the cost of not obtaining a sample balanced for the social variable under study, and the recording situation – while probably more naturalistic than task-based interactions, word lists, or interviews – further constrains participants' behaviour. Hence, we need to be careful not to generalise with respect to the exact distribution and frequency of the variables under study. However, this is not problematic because the central aim and contribution of this PhD thesis is theoretical and methodological, and the approaches presented here can be applied to bigger and better balanced samples of data in future studies.

3.2   Qualitative Methods

In this section, I first present a concise definition of Listener Responses as well as a summary of the *Action Types*, then outline some general coding decisions rooted in such a close analysis of the interaction, and finally present the transcription conventions I use here.

Listener Responses are defined as everything a Listener can do without challenging the Speaker for the floor *during* that Speaker's ongoing turn (most often a multi-unit-turn, but it can also be a shorter contribution). It is important to note that this definition is not based on form or length, but on sequential and interactional impact alone. A fuller definition is presented in chapter 4. Listeners can use this structural slot to do a number of different actions. A taxonomy of Action Types is developed and described in depth in chapter 5 and summarised in table 5.2. The Action Types coded for are:

1. acknowledgements

2. markers of surprise

3. first assessments

4. second assessments

5. self-initiated other-repair (other-completions of word-searches)

6. other-initiation (questions, requests for information)

7. joint constructions (discussed separately in chapter 5, but grouped together for the quantitative analysis)

(a) voicing

(b) collaborative completions

These seven categories are based on existing interactional and Conversation Analytic work presented in the Literature Review, as well as a close analysis of the over 5000 Listener Responses in the present set of recordings. The choice of these seven categories will be discussed in detail in chapter 5, where they are introduced and illustrated.

I will briefly summarise three sets of general coding decisions that are based on the Conversation Analytic approach to talk-in-interaction. It is essential to remember that all coding decisions are based on the sequential structure and the next-turn proof-procedure, i.e. the orientation the participants themselves display to the interaction. The three concerns to be discussed are multifunctionality, how to treat 'absences', and how to delimit individual Listener Responses and Action Types.

Firstly, multifunctionality is a common challenge in work on discourse markers (Pichler 2013a). Using the next-turn proof-procedure as a tool for coding makes it possible to develop and apply mutually exclusive categories. Neither of the two coders (see later section on inter coder reliability) noted any overlap at the level of Action Types. However, there is 'multifunctionality' if we consider different levels of the interaction: future work could for example investigate the relationship between alignment, affiliation, preference organisation and the individual Action Types. Nevertheless, all variants on any given level are mutually exclusive.[6]

Secondly, with respect to 'absences', Listener Responses are by definition optional, as noted in the literature review: They *can* be produced in specific places, but they very rarely *have to be* produced in any given place (see chapter 4).[7] Furthermore, they are often realised multimodally. Listeners can, for example, simultaneously nod, smile, or blink, *and* proffer a vocalised response. They can also only nod, blink, or smile (for different analyses of co-occurrence patterns see Hömke et al. (2017), Barth-Weingarten (2011), and Brunner (1979)). Accordingly, the absence of a vocalised Listener Response does not imply the absence of any response at all. This makes it impractical, and for the data at hand in fact impossible, to count 'absences'. Whenever I speak of Listener Response frequency here, I only mean the frequency of vocalised responses. Future studies could

---

[6]Of course in interaction these levels all work in concert and are not necessarily perceived as distinct. However, treating them separately in the analysis allows us to a) tackle the issue of multifunctionality, and b) investigate co-occurrence patterns once we have analysed the feature distributions on the separate levels.

[7]We could analyse individual deviant cases, in which a Speaker pursues a Listener Response. However, while this would illustrate the local relevance of that particular Listener Response in that particular interaction, it would still not be a practical strategy for counting 'absences'.

extend the methodological paradigm presented to include different multimodal cues and co-occurrence patterns.

The third set of concerns relates to identifying and delimiting Listener Responses and individual Action Types. As mentioned at the beginning of this section, all coding is based on the next-turn proof-procedure. This means that the specific realisation or form of an utterance is not part of the definition of each Action Type. The relationship between form and function is analysed in chapter 7. With this in mind, short paraphrases or repetitions of a Speaker's turn are coded as ACKNOWLEDGEMENT unless the Speaker orients to them as doing something else. When Listeners produce a number of Listener Response tokens in a row – repeating one token, or chaining different vocalisations – it can be challenging to delimit one Listener Response from the next. Following Stivers (2004) I treat repeated items or longer utterances as doing a single action if they are under one intonation contour, form one (Listener-)TCU, and are oriented to as one single action by the Speaker.

A related potential concern in talk-in-interaction are false starts and self-repair by the Listener. They are cued with hesitations, pauses, and separate intonation contours. False starts are extremely rare in Listener Responses and will not be coded separately. I follow the procedure common on sociolinguistic coding of only annotating the final, fully realised utterance (see Tagliamonte (2006: 94)).

The following transcription conventions, based on Jefferson (1984b) with small adaptations, are used to give an impressionistic representation of speed, amplitude, and pitch movement in the talk:

TRANSCRIPTION CONVENTIONS

quiet talk	°quiet°
quieter	°°very quiet talk°°
loud	<u>higher amplitude</u>
pitch rise	/up
strong rise	↑strongly up
pitch fall	\down
strong fall	↓strongly down

Turn-final intonation is indicated with punctuation at the end of each turn like this:

TURN-FINAL INTONATION

flat	this turn ends with no pitch movement–
slight rise	this ends on a slight rise,
strong rise	this one on a strong rise?
falling	and here it falls.

3.3 Quantitative Methods

After a brief word on the set of statistical tools used in the different parts of the analysis, I will present the results of the inter coder reliability tests.

The specific quantitative tools used vary slightly for each analysis chapter and will be presented in the context of the analyses. Generally, the descriptive statistics are based on group-internal percentages, and Zero-inflated Poisson regressions are used to model LISTENER RESPONSE frequency and variation in Action Types (chapters 4 and 6). Regression analyses are well-established tools in quantitative (socio)linguistic studies (see for example the discussion about how best to leverage their full potential in Drager and Hay (2012)), and Zero-inflated poisson regression is one of the less well-established sub-types, at least in sociolinguistics. Different kinds of poisson regression are used to model count data within and beyond linguistics, for example in psycholinguistics (Rigby et al. 2008; Lo and Andrews 2015), typology (Coupé 2018), or computationally advanced models of language change (Winter and Wieling 2016), and are very common in disciplines where the number of events of type x need to be modelled – for example incidence, fertility, or mortality in medicine (see for example Cleophas and Zwinderman (2018a), Gagnon et al. (2008), Jackson et al. (2016), and Mouatassim and Ezzahid (2012)).

3.3.1 Inter Coder Reliability

10% of each of the 24 dyadic conversations were pseudo-randomly selected for annotation by a second coder (Appendix G summarises which parts of each conversation were used for inter rater reliability (IRR), and how they are distributed over the recordings).[8] In total, nearly 500 Listener Responses were coded and annotated by both me and the second coder (see table 3.8). The second coder was Zac Boyd, at the time a fellow doctoral student in sociolinguistics with a focus on phonetic variation. He has no training in interactional analysis. This makes him a good representative of sociolinguists open to but not trained in interactional approaches, and allows me to evaluate how easy it would be to apply and implement this type of coding scheme more broadly. The coding manual with a summary of how the process of checking for inter-coder reliability was organised is given in Appendix H.

When discussing the coding process, the second coder emphasised how straightforward and easy to apply he found the coding scheme, both at the level of

---

[8]The common term used is IRR, based on inter *rater* reliability. I opt for the term *coder* rather than *rater* here, because the two people working on the data code for the presence of Listener Responses and the actions they do, rather than doing some numeric rating. I have, however, retained the abbreviation IRR because it is such a common term.

defining the phenomenon and with respect to assigning the Action Type categories. There were very few cases where he was unsure and checked with me while coding.

Given that the coding at hand involves two distinct processes, two different types of inter-coder reliability were calculated. First, the reliability of token identification will be discussed, based on F-measure, Recall (i.e identification of as many cases as possible), and Precision (i.e. correct identification of cases; no mis-identifications). The results show that my interaction-based definition of Listener Responses is accessible to and can be reproduced by sociolinguists with no specific interactional training. In a second, separate calculation, I show that overall there is high agreement between coders on the Action Type coding. This is a classic inter-coder reliability question we are familiar with from auditory coding. I use Krippendorff's Alpha (Krippendorff 2004a) for this. Results are presented in more detail in the next sections.

3.3.1.1   Identifying Listener Responses

When it comes to identifying LISTENER RESPONSES, there is no clearly defined number of answers – there are many possible ways of segmenting the conversation into turns, TCUs, or other parts of speech. Computational linguistic and machine learning work face a similar challenge, for example when Tweets (or segments in longer prose) containing a certain topic need to be identified by an algorithm (see for example Guzman et al. (2016) and Williams and Mahmoud (2017)). The reliability of these algorithms is assessed using a metric called *F-Measure*, which is the harmonic mean of *Precision* and *Recall* (Buckland and Gey 1994; Hripcsak and Rothschild 2005).

	Coder 1	
Coder 2	Listener Response	Talk
Listener Response	A	B
Talk	C	D

<div align="center">

TABLE 3.5
CROSSTAB TO ILLUSTRATE HOW TO CALCULATE PRECISION AND RECALL

</div>

Table 3.5 shows the four different aspects of the decisions the two coders make that go into the calculations of *Recall, Precision*, and *F-Measure*. The cell marked *A* represents those cases where both coders annotated a token as a LISTENER RESPONSE. Cell *B* is the count of cases in which Coder 1 annotated something as talk, and Coder 2 annotated the same instance as a LISTENER RESPONSE. Cell *C* contains the sum of the reverse cases: all those items that Coder 1 annotated as a LISTENER RESPONSE, but Coder 2 annotated as talk. Cell *D*, finally, contains the

count of cases that both coders agreed were talk, or in other words *not* LISTENER RESPONSES.

Recall (or sensitivity) tests if all cases that can be identified have been identified. It is a percentage based on the number of cases identified by the second coder relative to the all those identified by the first coder, who is treated as the reference value in the calculation. Incidentally, this is where a simple percentage agreement metric would stop and thereby fall short of the complexity of this coding process. *Recall* can be expressed as the following formula (letters refer to cells in table 3.5):

$$\text{Recall} = \frac{A}{A+C}$$

Precision, on the other hand, is a test used to ensure that the second coder has not only identified all possible cases because they indiscriminately marked up everything as a relevant case. It is a percentage based on the number of truly relevant cases identified by the second coder (i.e. cases marked as relevant by both coders) relative to the total number of cases identified as relevant by the second coder. This measure tests whether the coding decisions are accurate, or whether the second coder has over-identified the phenomenon. *Precision* can be expressed as the following formula:

$$\text{Precision} = \frac{A}{A+B}$$

The *F-Measure* combines these two values in one metric as the harmonic mean. It ranges from 0-1, and the higher it is the more reliable the coding is taken to be. Because *F-measure* combines *Precision* and *Recall*, it does not matter which coder is set as the reference value.

I will now present *Recall, Precision*, and *F-Measures* for the identification of Listener Responses in four steps. All calculations were done in R using the `caret` package. The results are summarised in table 3.6. In the following subsections, I briefly discuss each row.

	Precision	Recall	F-Measure
Initial	0.89	0.84	0.86
- Laughs	0.88	0.87	0.88
+ Coding changes	0.94	0.88	0.91
+ Discussion	0.99	0.96	0.97

TABLE 3.6
STEP-BY-STEP CALCULATION OF RECALL, PRECISION, AND F-MEASURE
FROM INITIAL CODING TO FINAL CODING AGREED ON BY BOTH CODERS.

Initial inter-coder reliability is high, with an *F-Measure* of .86, composed of *Precision* = .89 and *Recall* = .84. This level of *Recall* indicates that the second

coder identified 84% of the tokens marked as Listener Responses based on the interactional definition, while *Precision* (or correctness of identification) says that 89% of those tokens the second coder identified were actually Listener Responses – some irrelevant tokens were marked, but overall the coding is very reliable.

There were two systematic sources of disagreement, which will be treated separately: The first to be addressed is an oversight on the part of the second coder, who did not consistently treat laughter as a potential Listener Response, despite this being stated in the coding manual. The second has to do with a lack of explicit instruction in the coding manual on one specific class of sequences where tokens often used as Listener Responses are in fact *not* doing this action.

The first systemic disagreement with the second coder is addressed by removing all rows that contain laughter for both coders (reducing the number of annotations being compared by only 54, from 2173 to 2119). *F-Measure* increases slightly, most notable is a 3% increase in *Recall*, which is the number of Listener Responses identified out of all those that I marked up. However, *Precision*, or correctness of identification, decreases by 1%. This reflects the observation that the second coder did correctly annotate some instances of laughter as Listener Responses, but was not consistent in doing so.

In a second step, the issue caused by a lack of explicitness in the coding manual was resolved. The notions of incipient speakership and third turn receipts were not explained sufficiently clearly for a coder who is unfamiliar with interactional analysis.[9] Having been presented with these definitions and explanations, the second coder said he would not consider incipient speakership or third turn receipts Listener Responses any more, and found them easy to identify. All instances were discussed briefly, and then re-coded as *Talk* (i.e. not Listener Responses).

Based on this revised coding, *F-Measure* = .91, *Precision* = .94, and *Recall* = .88. This indicates that the second coder also identified 88% of all Listener Responses I marked, and that 95% of those tokens he marked as Listener Responses were also marked as relevant by me. The remaining disagreements were discussed, and almost all of them resolved.

They were primarily caused by cases at the fuzzy boundary between Speaking and Listening. Some examples are discussed in the analysis chapter outlining the definition of Listener Responses and the individual Action Types. The *F-Measure* results for the final agreed coding were *F-Measure*= .97, with *Precision* = .99, and *Recall* = .96.

Overall, IRR for the identification of Listener Responses was extremely high, with an initial *F-Measure* of .86, and a final *F-Measure* of .97. This is corroborated by

---

[9]Some of these cases will be discussed in chapter 5 because they relate to the crucial question of delimiting the variable.

the second coder's comment that the identification was straightforward and very rarely problematic. It compares favourably to Duncan and Fiske (1977: 49-50), who report Intraclass correlations between coders as their inter coder agreement. They find Intraclass correlation of .90 for short backchannels, .88 for long backchannels, and .95 for all backchannels. When it comes to token identification, my interaction-based definition of LISTENER RESPONES is at least as reproducible as previous, less detailed definitions.

The cases of initial disagreement between coders were partly based on human factors, partly on differences in training background (something that can be addressed by more training and a more explicit coding manual), and a marginal number relates to challenge of imposing a clear border between Speaking and Listening. The remaining low level of disagreement serves as a reminder that this boundary is not always clear-cut, and that interactants share and continuously co-construct the interactional space.

3.3.1.2 Action Type Annotation Agreement

Krippendorff's alpha (Krippendorff 2004b; Krippendorff 2004a; Krippendorff 2011) was used to calculate inter coder reliability for the Action Type coding. This metric assesses how likely the *dis*agreements between coders are due to chance. This makes it a more conservative measure than, for example, Cohen's kappa (Cohen 1960) or other measures focused on agreement between coders.[10] α can range from 0 to 1, with 0 indicating 'perfect' disagreement, and 1 indicating perfect agreement. Krippendorff (2004a: 241) suggests $\alpha > .800$ as the cut-off point at which it is reasonably safe to say that agreement between coders is not due to chance. $\alpha > .667$ is presented as the lowest acceptable value that indicates broadly reliable agreement.

Table 3.7 presents an overview of the three ways I have calculated Krippendorff's alpha: The first value is based on a binary distinction between *acknowledgements*, which account for approximately 80% of the tokens, and the remaining *other Action Types*. The second α is based coding for all Action Types, and the third on the 20% that were assigned the six different *other Action Types* (for an overview of the *other Action Types*, see 5.2 in chapter 5). The analysis presented here is based on the coding excluding laughter tokens.

The first crucial distinction a coder needs to make is that between *acknowledgements* and all the other Action Types. The two coders had very high agreement, with $\alpha = .85$, above the threshold Krippendorff (2004a: 241) recommends for treat-

---

[10]For a discussion why Krippendorff's alpha might be deemed more reliable than previous metrics see Hayes and Krippendorff (2007)

	Krippendoff's Alpha
Binary	0.85
All Action Types	0.73
'Other' Actions Only	0.50

ing the agreement as reliable. Because this is a binary decision, we can compare the Alpha value to *F-Measure*, setting *acknowledgement* as the 'relevant' coding choice. The results were comparable to the token identification task **after** discussion, with *F-Measure* = .97, *Recall* = .98, and *Precision* = .96. This indicates that the main and crucial distinction between *acknowledgements* and the six more complex Action Types is intersubjectively clear, and reproducible by different coders.

When Krippendorff's Alpha is calculated for all seven Action Types, agreement is slightly lower, with α = .74. Inter-coder agreement can be expected to decrease as the number of categories increases. Nevertheless, the overall agreement between both coders is still reliable according to Krippendorff's suggested thresholds.

Given that these six 'other' categories are clustered in only 20% of the coded tokens, I calculated Krippendorff's Alpha separately for this sub-group, resulting in α=.50. This is fairly low compared to the overall high rates of agreement between the two coders, but not surprising. Firstly, as complexity in decision-making increases, so does the chance of a disagreement. Secondly, some of the distinctions between Action Types are more difficult than others. Table 3.8 shows a count-based confusion matrix of the two coders' coding decisions, and the heat map in figure 3.1 visualises the coding as proportional agreement.

The ratios given in the heat map (3.1) treat my coding as the target, and compare how Zac coded the tokens in each category as identified by me. Thus, each column sums to 1 (i.e. 100 %) from top to bottom. The colour scale ranges from light to dark, with light indicating high and dark low ratios. For clarity, proportions are given for each facet within the plot. The final (post-discussion) agreement of Zac's with my coding of *acknowledgements* is represented in the bottom left corner. Out of the 399 *acknowledgements* I identified, Zac also identified 390 or 98% as such. He identified 1% as a *joint construction* and 1% as a surprise mark. Table 3.8 contains the raw token counts to the proportions 3.1.

Coder 1

Coder 2	acknowledge-ment	1st assessment	joint construction	surprise mark	other-initiation	2nd assessment	self-initiation	Total
acknowledge-ment	390	5	1	3	0	6	1	406
1st assessment	1	9	1	1	0	2	0	14
joint construction	4	1	17	0	2	1	0	25
surprise mark	2	3	0	15	1	0	0	21
other-initiation	1	0	0	0	11	0	0	12
2nd assessment	0	2	0	0	0	4	0	6
self-initiation	1	0	7	0	0	0	7	15
Total	399	20	26	19	14	13	8	499

TABLE 3.8
CONFUSION MATRIX WITH COUNTS FOR THE TWO CODERS.

77
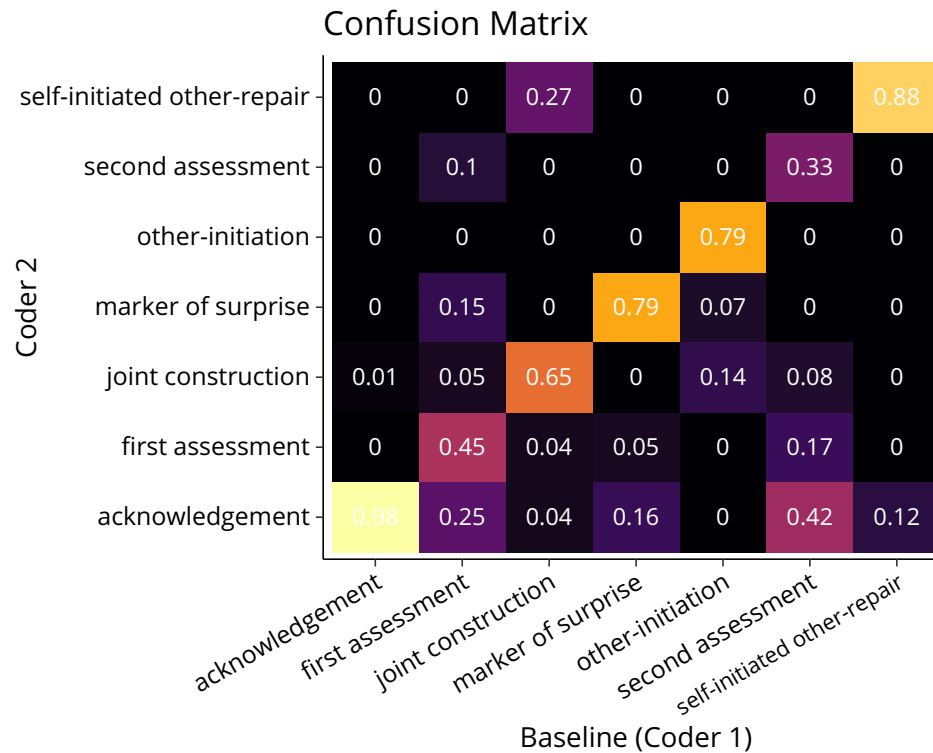
FIGURE 3.1
CONFUSION MATRIX COMPARING THE TWO CODERS' DECISIONS BY
ACTION TYPE CATEGORIES.

This Action Type-by-Action Type visualisation of inter coder agreement shows high agreement in coding decisions overall; the low α for *other Action Types* is driven by two, maybe three categories: There is extremely high agreement between the coders on *acknowledgements* (98%) and *self-initiated other-repair* (88%), both coloured in white-yellow. Agreement on *surprise marks*, *other-initiations* (79% for each) is also very high. In contrast to this, agreement on *joint constructions* (65%) is relatively low, and that on *first* (45%) and *second assessments* (33%) below chance.

It is important to keep in mind that there are only 8 to 26 instances of Action Types beyond *acknowledgements* (see table 3.8), which makes the results presented indicative only. Nevertheless, they do reflect qualitative considerations related to the Action Type coding scheme very well.

First and *second assessments* frequently get mistaken for each other, putting into question the utility of separating them. I will discuss this when presenting the different Action Types in Chapter 5. Both kinds of *assessments* are relatively frequently taken for *acknowledgements*, especially if their lexical form is *yeah*. In these cases, Action Types can only be distinguished based on the sequential structure (which would require more specific training or a background in interactional analysis) and the prosodic realisation of the lexical item. The same issue underlies

the confusion between *first assessments* and *surprise marks*, and *surprise marks* and *acknowledgements*, as well as that between *joint constructions* and *self-initiated other-repair*. The complex form-function relationship between the different actions and the lexical material used to do them will be discussed in Chapter 6.

Overall, IRR by Action Type shows high agreement, with the low α for actions beyond *acknowledgements* largely driven by the interactionally most complex types, especially those where the same lexical material can be used to do different actions.

3.3.1.3   Summary of all IRR results

The two sets of IRR presented above demonstrate that both the definition of Listener Responses and the *Action Types* used in this study are reliable and inter-subjectively reproducible. We have successfully identified Listener Responses, and within those annotated which are *acknowledgements*, by far the most frequent *Action Type*. Agreement decreases somewhat as the number of *Action Types* and thereby options increases. The majority of disagreements, both in token identification and *Action Type* coding, are driven by boundary-cases and serve as a reminder that the complexity of interaction cannot be fully captured by clear-cut coding categories (on this, see also Schegloff 1993).

Agreement between coders is very high, particularly considering that the second coder had no previous training in interactional or Conversation Analysis. Similar studies could expect to reach even higher agreement by providing a more highly specified coding manual, better training, and/ or coders who have a background in interactional analysis. Based on the analysis of disagreements, a revised coding manual as well as any training should include a very explicit introduction to the next-turn proof-procedure, a discussion of incipient speakership and third turn receipts, and exemplary analyses of relevant examples.

Part III

Analysis

Chapter 4

How often do Listeners Respond? Defining the Variable and Operationalising Frequency

> What stands fast does so, not because it is intrinsically obvious or convincing; it is rather held fast by what lies around it.
>
> *Ludwig Wittgenstein* (On Certainty: 144)

To date, there is no sociolinguistic approach that allows us to quantify LISTENER RESPONSE frequency in a way that is sensitive to interactional structure at the turn-level. In this chapter I propose operationalising frequency based on the number of words in the turn that is being responded to. This enables us to analyse variation, in this case by Speaker and Listener gender. The methodology presented here also invites us to reflect on the assumption that the individual Speaker or the Speech community is the locus (and envelope) of variation. Variables like LISTENER RESPONSES are rooted in the interactional structure and need to be considered within the system of the interaction. This means variation is situated not within one individual but in the shared space created by the co-participants in an interaction. Consequently, we need to quantify not based on variables intrinsic to the individual, but variables pertinent to the interaction – in this case, the length of the turn that is being responded to.

Previous work on variation in LISTENER RESPONSE frequency has looked at inter-group differences with respect to culture (White 1989; Heinz 2003; O'Keeffe and Adolphs 2008) and gender (Kogure 2003; Reid 1995; Fellegy 1995). While both vary to some extent in the corpus under analysis here, gender differences are better represented and therefore make for a better social variable to demonstrate what can be done with the proposed methodology.Crucially, the contributions of

this dissertation are theoretical and methodological; gender is simply used as a case-study. The methodology presented here can be applied to any other social variable, for example age, ethnicity, or other more locally constructed categories.

Quantifying overall frequency is the broadest perspective and first step in developing an interaction-based approach to LISTENER RESPONSES as a sociolinguistic variable. It takes into account the distribution of Listener and Speaker roles. I treat 'Speaker' and 'Listener' as roles participants swap in an interaction. For the Speaker to remain the Speaker, the Listener has to reaffirm their Listenership from time to time. As noted in the introduction, I will be referring to the turn-holder as *Speaker* with a capital S and to the person doing listening/producing a LISTENER RESPONSE as *Listener* with a capital L.[1]

The chapter is structured as follows: I first develop a definition of LISTENER RESPONSES as a sociolinguistic variable based on a close analysis of conversational extracts, using the next-turn proof-procedure. I describe the variants, i.e. a LISTENER RESPONSE being present or absent, and the envelope of variation. Based on this, I propose an interactionally accountable way of quantifying frequency. The main results are presented through descriptive and inferential statistics. The descriptive analysis shows cross-gender accommodation in terms of how often men and women produce LISTENER RESPONSES when listening to a male or female Speaker. LISTENER RESPONSES are the most frequent in all-female and the least frequent in all-male dyads. The inferential statistics, a Zero-inflated Poisson regression, confirms this trend, but also reveals that turn length is the only statistically significant effect in both the logit and the count model, while the role of the social variables is more complicated. Crucially, the gender of the person *receiving* the response matters statistically more (marginally significant at p= 0.048) than that of the person *producing* the responses (ns).

I then discuss the implications of integrating interactional and variationist methods, particularly the impact of the next-turn proof-procedure and the change in perspective brought about by considering the interaction as the envelope of variation. I illustrate the importance of this by contrasting the quantification proposed here with three previous ways of quantifying LISTENER RESPONSES.

---

[1]The expression '*doing* Listening' or '*doing being* a Listener' is used here to underline that these are processes the participants are actively involved in negotiating at any given time rather than fixed states.

4.1 Analysis

The analysis is structured as follows: I first define LISTENER RESPONSES as a variable, describing it based on its sequential position, and outline the envelope of variation. I then present descriptive statistics looking into variation between male and female Listeners and Speakers. Because there is large within-group variation, I move on to present a Zero-inflated Poisson regression model with Listener and dyad as random effects.

4.1.1 Defining The Variable

LISTENER RESPONSES are defined based on their sequential position and interactional impact: whether a contribution made by any participant is a LISTENER RESPONSE is determined using the next-turn proof-procedure.[2] In essence, the next-turn proof-procedure is based on the understanding that we can only infer the interactional meaning of a contribution based on how it gets treated in the next turn. For LISTENER RESPONSES this means if a Speaker produces a (usually longer) stretch of talk, called a multi-unit-turn, and the Listener responds in some way, we need to check how the Speaker treats this response. If they continue their ongoing project, the Speaker has oriented to the Listener's contribution as a LISTENER RESPONSE. Crucially, this reflects the understanding that interaction is always collaborative and co-constructed, and the roles of Speaker and Listener are constantly (re-)negotiated.

It also implies that when one person is Speaking, the Listener is not passive at all. On the contrary, they play an important and active part in the other's Speaking in two complementary ways: firstly, by *not* producing talk that would claim the floor while the other holds it, and secondly by producing vocalised and multimodal supportive feedback. This active affirmation of the Speaker-Listener role-distribution is what I refer to as LISTENER RESPONSES here. They have three key characteristics:

1. As described above, they orient to the current Speaker-Listener role distribution in three different ways:

   (a) They do not claim the floor; this is signalled through their brevity, their prosodic design, and/or because they elicit more Speaker-talk.

   (b) They can be produced after a short lapse, and usually the Speaker then picks up the thread.

---

[2]For an in-depth discussion of this Conversation Analytic tool, see the literature review and chapter 6 within this thesis, and for example Atkinson and Heritage (1984), Sidnell (2012), and Clift (2016)

(c) If there is a lapse following the Listener Response, the Listener generally waits until the Speaker continues rather than taking over the floor. If they have waited and signalled in other ways that they expect the Speaker to continue, the Listener may eventually take the floor.

2. LISTENER RESPONSES *tend* to be placed at places of syntactic, prosodic, and pragmatic completion (so called Transition Relevance Places, TRPs; see Ford and Thompson (1996) and Clancy et al. (1996))

3. They are mostly optional – there are certain places of *possible* occurrence, but it is not necessary for the Listener to provide a response at *every* point where one might be relevant.

In brief: LISTENER RESPONSES are all those things Listeners can do in the Speaker's interactional space without disrupting the Speaker's ongoing turn at talk. This is represented schematically in the structural pattern given below, and I will elaborate this basic definition of the variable based on several examples from extract 1.

STRUCTURAL DEFINITION OF A LISTENER RESPONSE

```
1    Speaker:    ongoing multi-unit-turn
2    Listener:   Listener Response
3    Speaker:    continuation of ongoing multi-unit-turn
```

Extract 1 is taken from a conversation between Donna, a social worker in her early twenties who was diagnosed with diabetes as a pre-teen, and PuzzleB, whose teenage daughter has Type 1 Diabetes. It starts with a story about PuzzleB's daughter's experience of attending a diabetes camp (lines 1-33), then Donna offers a brief sequence about her experience at a DAFNE course (lines 34-47), and then PuzzleB reclaims the floor to expand on her narrative about the diabetes camp (lines 48-end). These two floor changes show that Speakership and Listenership are not assigned for a fixed amount of time, but that they are continuously (re-)negotiated. It is also worth noting that the respective Listener does a variety of actions throughout the other's multi-unit-turn – this diversity within the broader category of LISTENER RESPONSES will be the focus of analysis chapters 5 and 6. Furthermore, sometimes those tokens conventionally used to signal Listenership are used to take the floor. This is called signalling 'incipient speakership', i.e. signalling that someone who has been the Listener is preparing to take the floor. It is important to differentiate these from LISTENER RESPONSES, and I will discuss some examples from the extract below to illustrate this point. LISTENER RESPONSES are marked in plum and instances of incipient speakership in orange.

```
1      PuzzleB:    so we put her on the train and he he-
2      PuzzleB:    she was she was texting me and she said she must've
3                  been in-
4      PuzzleB:    I don't know-
5      PuzzleB:    Carlisle- not Carlisle um somewhere on the borders-
6      PuzzleB:    Lockerbie it was just-
7      PuzzleB:    'My °blood's°< <laughing> twenty- [↑fou:r,'>
8      Donna:                                        [ha
9      PuzzleB:    I'm < <laughing> 'Oh my /go:d,'>
10     Donna:      [< <laughing> °come /ba:ck°->
11     PuzzleB:    [< <laughing>I think the stress was kicking in=
12     PuzzleB:    =he he he[he he->
13     Donna:               [he he he-
14     PuzzleB:    and then there was no reception after that for a while-
15     PuzzleB:    anyway she had the most- well because she met people
16                 she said that just understood her.
17     PuzzleB:    and understood the condition-
18     PuzzleB:    and I think a lot of them- they were f- all from
19                 England,
20     PuzzleB:    but I think a lot of them'd been sent-
21     Donna:      °mhm°.
22     PuzzleB:    because they weren't controlling- (.)
23     PuzzleB:    well she controls her (.) type one I think she's had
24                 excellent healthcare in Glasgow.
25     PuzzleB:    but I think a lot of them (.)-
26     PuzzleB:    were (.) [just not looking [after themselves.
27     Donna:               [°°weren't°°      [°yeah°-
28     PuzzleB:    but she still enjoyed it just that (.) [relaxation you
29                 know=
30     Donna:                                             [°that's good°.
31     PuzzleB:    =it's just like 'Oh, being with other people who
32                 understand this and-'
33     PuzzleB:    (.5)
34     Donna:      yeah cause I'd never-
35     Donna:      well I'd met people-
36     Donna:      but I never knew anybody-
```

85

```
37      Donna:        that was diabetic until I did that DAFNE course,
38      PuzzleB:      /right oka[y\.
39      Donna:                  [um::-
40      Donna:        and so it was good meeting them-
41                    (.6)
42      PuzzleB:      [right,
43      Donna:        [just and we're all in a whatsapp group now-
44      Donna:        and ju[st having people that-
45      PuzzleB:            [↑oh really\.
46      PuzzleB:      er:::-
47      Donna:        understand,
48      PuzzleB:      and how long ago did you go on that course,
49      Donna:        [!] um:: two years ag[o-
50      PuzzleB:                            [two years ago.
51      PuzzleB:      ↑oh right sounds quite good isn't it that there's still
52                    the:-
53      Donna:        yeah:[:-
54      PuzzleB:           [yeah and the-
55      PuzzleB:      yeah cause that's what happened my daughter after going
56                    on the: um camp is that-

              PUZZLEB CONTINUES
```

In the first six lines, PuzzleB unfolds the story of her daughter on her way to the diabetes camp, and Donna produces the first acknowledgement on line 7; a burst of laughter in overlap with the end of PuzzleB's turn. PuzzleB continues without making any audible reference to this Listener Response; she simply reports her response to her daughter's message: oh my god (line 9) bracketed by laughter. Donna, engaging in the narrative, produces a possible continuation of PuzzleB's turn: come back (line 10), also shaded with laughter. PuzzleB continues in overlap with Donna's Listener Response, giving a possible explanation for her daughter's hyperglycaemia – the excitement and nerves related to the diabetes camp. As soon as Donna has finished her possible continuation of PuzzleB's turn she moves back into laughter (lines 13/14), laughing along with PuzzleB's continuation and signalling her continued listenership. Both speakers orient to Donna's continuation not as a speakership bid but as a LISTENER RESPONSE that reinforces rather than questions their current Speaker-Listener relationship. In all the examples so far, the LISTENER RESPONSE has been produced in overlap with the ongoing talk, and the Speaker has simply continued with their ongoing project.

From lines 15 to 20 PuzzleB develops the story without further vocalised responses from Donna (though it is likely that Donna signals her continued listenership through cues at the level of gaze, posture, and facial expression). Line 20 can be understood as projecting further talk, given the flat intonation and the juxtaposition of her daughter wanting to go to the camp (which PuzzleB had mentioned at the story launch) with the English participants having been sent. Donna produces the LISTENER RESPONSE mhm (line 21) in the brief pause between PuzzleB's turn constructional units (TCUs). PuzzleB's continuation makes no direct reference to the mhm; it has not disrupted the ongoing talk in any way.

The other participants' diabetes management remains a topic of concern for PuzzleB and she compares her daughter with the other adolescents in the camp several times. Up to this point her speech has been very quick and fluent, but she briefly hesitates as she makes these potentially face-threatening statements. In line 27, Donna produces another Listener Response, weren't, in overlap with PuzzleB's talk. This is a possible completion of PuzzleB's turn, and when Donna hears that this is exactly how PuzzleB continues her turn, she proffers the more prototypical acknowledgement yeah in overlap with the continuation. Both these LISTENER RESPONSES are quieter than the surrounding talk and are not laying any claim to the floor.

PuzzleB moves away from these critical comments in line 28, saying her daughter still enjoyed the camp, and Donna responds with a first assessment in line 30. This first assessment (see Pomerantz (1984) on assessments) overlaps with PuzzleB's ongoing turn at a point of syntactic, semantic, and prosodic incompletion. Continuation of the turn is hence highly projectable. The assessment follows a very brief pause within PuzzleB's turn, but it is not solicited, neither lexically through a tag question nor prosodically through rising intonation or a marked pause. The assessment is therefore optional, while it is simultaneously designed to specifically respond to the preceding talk in treating something in that talk as an assessable. It is lexically minimal and prosodically reduced, two additional cues that it is not an attempt to claim the floor.

PuzzleB pauses for .5 seconds (line 33) once her turn has reached a point of possible completion. This is the longest silence in the conversation so far, and Donna reaffirms her orientation to the current role distribution by waiting for a full .5 seconds before she takes the floor in line 34, using the acknowledgement token yeah to initiate her turn and talk about her own experience of meeting other people with Type 1 Diabetes (see Jefferson (1993) and Jefferson (1984a) on the use of 'yeah' to preface immediate on-topic continuations).

Donna's statement that for her this first experience was a DAFNE course in Glasgow is syntactically but not prosodically complete at the end of line 37.

PuzzleB responds with an acknowledgement `right okay`, also marked with rise-falling intonation, which makes further speaker-talk relevant. Donna does indeed continue in overlap with the last part of the final vowel of the `okay` – albeit initially with a hesitation marker only, followed by a summary assessment (line 40). Both participants here orient to the current Speaker-Listener roles and to the relevance of Donna continuing her talk: She occupies the floor and signals that there is more talk to come with the `um::` (line 39) instead of leaving a silent pause.

There is a .6 second pause following Donna's summary assessment, and PuzzleB responds with the acknowledgement `right` (line 42). This acknowledgement has rising intonation again, making further Speaker-talk relevant. In fact, Donna continues with more on-topic-talk in overlap with the acknowledgement. She mentions that the DAFNE participants have created a WhatsApp group and have been keeping in touch since the course. PuzzleB treats this as new and surprising information, responding with a surprise mark (line 45) in overlap with Donna's TCU.

PuzzleB then uses the vocalisation `err` (line 46) to signal her incipient speaker-ship, and uses a request for information (line 48) to reclaim the floor once Donna has finished her turn in line 47. PuzzleB again places the tokens `oh right` and `yeah` in lines 51, 54, and 55 at the start of the respective TCUs, receipting Donna's talk but at the same time signalling the change in Speakership. Donna could contest PuzzleB's claim to the floor by expanding on her answer in line 49, but she moves into Listenership with an acknowledgement in line 53, in response to PuzzleB's assessment of her DAFNE experience. PuzzleB now uses the very same lexical tokens that have previously served her as LISTENER RESPONSES to signal receipt of Donna's talk. In contrast to their use as LISTENER RESPONSES, such tokens tend to be louder and quicker when they signal incipient speakership, and their sequential position is different – signals of incipient speakership are followed by more talk from the same person, while LISTENER RESPONSES are followed by more talk from the person who has been holding the floor.

I point this out to demonstrate the importance of defining LISTENER RESPONSES based on their sequential and structural properties and not on form alone – a form-based definition might treat these turn-initial uses of `oh right` and `yeah` as equivalent to their use where they (re)affirm listenership (Drummond and Hopper 1993a), and a study looking at all the uses of `yeah` would miss out on all the other LISTENER RESPONSES we have already seen.

These observations build up to showing that LISTENER RESPONSES are a truly interactional phenomenon. The listening slot is not a neutral space but rather one constantly co-created by the two participants. To summarise, LISTENER RESPONSES are defined as follows:

1. They support the current Speaker in their talk by being unobtrusive, brief, or eliciting more same-Speaker talk.

2. They *tend* to be placed at TRPs, **but**

3. they do not *have to be* produced at every possible point of occurrence, and their occurrence is not restricted to TRPs.

The very high rate of inter coder reliability for annotating Listener Responses confirms that this definition is intersubjectively tenable (*F-Measure* = 0.97, *Recall* = 0.96, *Precision* = 0.99, see section on inter coder reliability in chapter 3)

Based on the definition presented here, it is to be noted that LISTENER RE-SPONSES are firmly rooted in the interactional structure, and we need to take this into account when operationalising frequency. We have also already seen that even though all LISTENER RESPONSES share the same overall structural property of supporting rather than disrupting the ongoing Speaker talk, they can do a number of different actions. A strong orientation to the interactional context and the use of the next-turn proof-procedure as an analytical tool will therefore become even more crucial once we analyse variation in these different actions Listeners can do (chapters 5 and 6).

4.1.2 Quantifying LISTENER RESPONSES turn by turn

Now that we have an interactional definition of LISTENER RESPONSES, we need to develop an interaction-based way of quantifying them. As pointed out above, LISTENER RESPONSES *tend* to be placed close to transition relevance points, i.e. the edges between intonational, syntactic and pragmatic units (Ford and Thompson 1996; Clancy et al. 1996). This might suggest TRPs are the envelope of variation. However, LISTENER RESPONSES do not *have* to be produced at *every* TRP, and they can also occur in overlap with an ongoing turn constructional unit (TCU). This means treating TRPs as the envelope of variation does not actually include all possible contexts of occurrence. It also means we cannot count 'absences' of LISTENER RESPONSES, only their presence.

Considering that Listeners orient to what is happening in the ongoing turn-at-talk, they can be seen as responding to the turn as a joint project under construction rather than to each TCU individually. In fact, they need to orient to both these levels simultaneously, and they do. Here, we focus on the broad level of producing vs not producing a LISTENER RESPONSE, which attends to the general role-distribution in that stretch of interaction. Listeners also orient to the immediate sequential context by producing *different kinds of* LISTENER RESPONSES – which action they do is contingent on the preceding TCU, and LISTENER RESPONSES can be

taken up in different ways by the current Speaker. As mentioned previously, this will be the focus of the next two analysis chapters; this first introductory chapter outlines a big-picture approach to Listener Response frequency. Hence, the best approximation to conceptualising the envelope of variation is the length of the turn that is being responded to, measured in number of words in that turn.[3] I will return to the theoretical and methodological implications of this approach in more detail in the discussion. A 'turn' is operationalised as all the talk tagged as 'talk' by one Speaker, until the other participant produces something tagged as 'talk' in the transcript. Given that the definition of 'turn' is notoriously challenging, this could certainly be further refined. However, based on the extremely high inter rater reliability of annotating Listener Responses, this operationalisation of 'turn' based on changes of 'Listener' and 'Speaker' status in the annotated transcript is justifiable.

Extract 1 just presented exemplifies such a floor change (and the challenge in defining the exact cut-off of a turn): from lines 1-32, PuzzleB holds the floor and Donna produces Listener Responses. Then there is a lapse in line 33, and Donna takes over the floor with a turn that starts with a prototypical Listener Response token which leads into the new turn. The automatic word-counter counts all of the preceding words produced by PuzzleB as 'turn-words', skipping Donna's words that are marked as Listener Responses, and then starts a new 'turn-word count' for Donna as soon as her utterance is marked as 'Speaker-words' or 'not a Listener Response'. We see that Donna's turn here goes until line 47 – up to this point, PuzzleB's contributions are marked as Listener Responses and thus the counter continues to add all of Donna's word to her turn word-count. In line 48, PuzzleB asks a question and receipts Donna's answer (given in line 49) with a 3rd turn receipt (in line 50). Therefore, PuzzleB's turn in line 48 is *not* marked as a Listener Response, and this line ends the turn word-count for Donna and starts a new turn and turn word-count for PuzzleB.

---

[3]Some readers might be concerned because I am relating the number of Listener Responses (which can consist of more than one word each) to the number of words in the ongoing turn. This might look like comparing apples and oranges. However, because the variable I am conceptualising here is Listener Responses, the unit of analysis is the Response, not the number of words in the response. Further, the vast majority of Listener Responses are in fact very brief (one word only), but some particular action types can be longer. Relating number of words to number of words would overstate the importance of those actions relative to the rest. Hence, it is not a problem but rather a methodological necessity to relate number of Listener Responses in a turn to the number of words in that turn.

The formula used to calculate Listener Response Frequency that I have just described is given below:

$$\frac{\text{Number of Listener Responses reacting to turn X}}{\text{Number of words in turn X}}$$

This rate is calculated for each individual turn, and then multiplied with 100, the mean turn length, in order to make it possible to compare turns of different lengths. This metric is a refined version of that proposed by Duncan and Fiske (1977), who calculated number of Listener Responses per 100 words in the overall conversation rather than for each turn individually. Previous work using different metrics has mostly found no gender difference in overall frequency but often note differences in placement or type of response. Dixon and Foster (1998) who apply the metric proposed by Duncan and Fiske (1977) note that (1) context supersedes any gender effect, with participants producing fewer responses in competitive than in non-competitive conditions, and (2) they find an interaction between Listener and Speaker gender: when listening to a female Speaker, male Listeners produced *more* responses than female Listeners. They hypothesise this could be explained by male participants wanting to win the approval of the female Speakers and (over)accommodating to them. Further, they note that this puts into question the idea of the 'unsupportive' male Listener, and context playing a much bigger role than gender does not support the idea that there is a 'men's/women's language'. Those interpretations thus question both a 'dominance' approach to gender (women as being more supportive, more facilitative than men because of their lower social status and power), as well as a 'difference' approach that assumes underlying innate differences. To fully understand what these behaviours and differences in overall frequency of responding mean interactionally, we need to analyse them in context. It is entirely possible that the overall similarity in frequency masks an underlying difference in *what* Listeners do, and *how* they do it. This analysis is left for the rest of the thesis – for now we shall look into how the Listeners in my study behave with respect to overall Listener Response frequency.

4.1.3   Descriptive Statistics

The data consist of 24 dyadic, topic-focused conversations between two participants each. Both participants take the role of the Speaker at some points, and that of the Listener at others. This means that each dyad contains four roles: each participant as the Speaker, and each participant as the Listener respectively. Consequently, in 24 conversations with two participants each, this results in 48

'Speaker-roles' and 48 'Listener-roles'. There are 10 all-female dyads, 13 mixed dyads, and 1 male-only dyad. Given that each participant takes both the Speaker and the Listener role at some point in the interaction, this means there are 20 female-female Listeners and Speakers, 13 female-male and male-female Listeners and Speakers, and only 2 male-male Listeners and Speakers.

Thus, there is a fair amount of data for all-female and mixed-gender dyads, but data for all-male dyads is very sparse. This is the case because the data were collected in the context of a bigger project investigating how people living with Type 1 Diabetes in Scotland talk about their condition and healthcare provision, and balancing the sample for social variables like gender was not the primary aim. This is not an issue for the analysis presented here, because my goal is to make a methodological point, not one about gender as such. However, I would like to caution the reader once more to be particularly careful not to generalise from the one all-male dyad. Table 4.1 gives an overview of how many LISTENER RESPONSES are produced by male and female Listeners towards male and female Speakers.

Listener \ Speaker	Female	Male	Total by gender
Female	2291	1557	3848
Male	1100	254	1354

TABLE 4.1
CROSS-TABULATION OF RAW TOKEN COUNTS OF LISTENER RESPONSES BY
LISTENER AND SPEAKER GENDER ACROSS ALL INTERACTIONS

Overall, there are 5202 LISTENER RESPONSES, 3848 of which are produced by women, and 1354 by men. If we relate this to the gender distribution in the sample, this is not surprising: 33 (or 68%) of the 'Speaker slots' are occupied by women, and only 15 (or 32%) by men, and an analysis of 'Speaker-words' vs 'Listener Responses' for all male and female participants shows an even distribution of Speaker- and Listener-roles. This suggests that male and female participants behave similarly with respect to how frequently they produce LISTENER RESPONSES (for the raw counts of Speaker-words and LISTENER RESPONSES see appendix I).

The more talk a Speaker produces in one multi-unit-turn, the more opportunities a Listener has to respond. I will explore the nature of this relationship in the following sections. First, let us get an overview of the data structure and see how turn length is distributed in the conversations (figure 4.1).

Figure 4.1 shows that turn length ranges from one to 570 words, with a mean (marked with the dotted vertical line) close to 100 words. It further visualizes that turn length follows a log-normal distribution. In the following analyses, turn length will therefore be log-normalised. The distribution of number of LISTENER RESPONSES
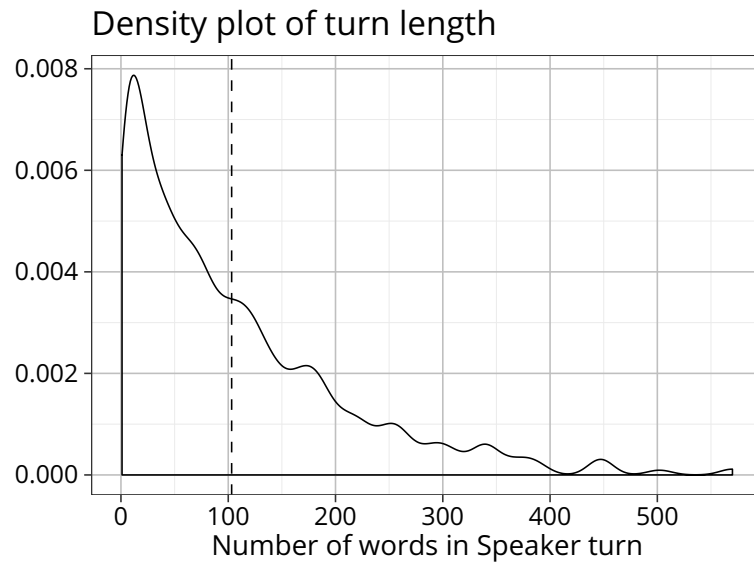
Figure 4.1
Density plot showing turn length follows a log-normal distribution, with the mean being at 103 and the median at 75. Turn length ranges from 1 to 570 words.

per turn follows a similar highly positively skewed distribution, with the majority of values at 0. This means, for most turns of length x, there are 0 Listener Responses. To reduce the skewness, Listener Response frequency per turn was log normalised with `logp1()` in R (see Baayen (2008)). The `log1p()` function in R calculates log(x + 1) which is necessary for Listener Responses, where the value can be 0. The visualisations of observed and predicted number of Listener Responses relative to turn length will present the raw numbers, with the axes using the log-scales. This combines the benefit of intuitively understandable numbers and a visualisation appropriate to the nature of the data.

Based on what I have outlined above, the length of the ongoing Speaker-turn is the linguistic factor that is most likely to influence the number of Listener Responses – as turn length increases, number of Listener Responses is likely to increase. Because interaction is constantly co-constructed by both parties involved, it is imperative from a theoretical point of view to consider not only the social characteristics of the Listener but also those of the Speaker – here their gender. Previous work has found an effect of Speaker gender (Reid 1995; Kogure 2003).

Figure 4.2 shows number of words in the Speaker-turn on the x-axis, and number of Listener Responses on the y-axis. The axes are log-scaled to account for the log-normal distribution of the two variables. The plot is faceted by Listener gender and has separate lines to represent Speaker gender; solid for female and dotted for male Speakers. The lines were created with ggplot2's `geom_smooth()`

Listener responses by turn length
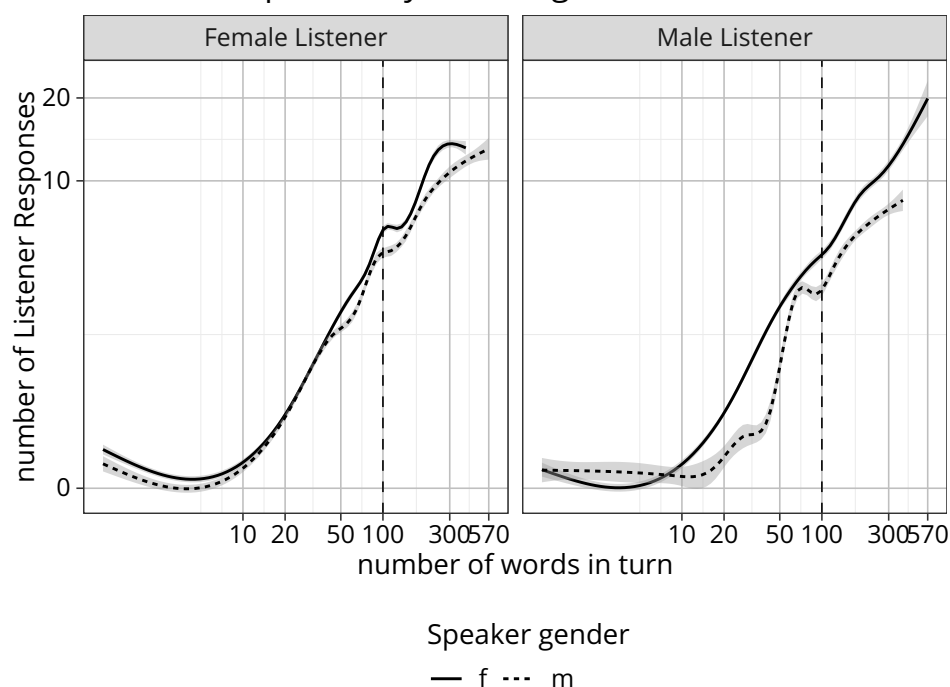
FIGURE 4.2

NUMBER OF LISTENER RESPONSES INCREASES WITH TURN LENGTH, BUT
TO A DIFFERENT EXTENT DEPENDING ON SPEAKER AND LISTENER GENDER.
FEMALE LISTENERS PRODUCE LISTENER RESPONSES MORE FREQUENTLY
THAN MEN.

function and simply serve to illustrate the trend. They are not equivalent to the Zero-inflated Poisson regression model presented in the next section on inferential statistics. The left facet shows how female Listeners' number of responses increases with Speaker turn length for both female and male Speakers. The right facet shows the same for male Listeners. We can see that, a few outliers aside, Listeners respond to turns that are 10 words or longer, and the number of responses increases as the number of words in the turn increases. We also note a difference between male and female Listeners: generally, women are shown to produce LISTENER RESPONSES than men, especially when listening to a female Speaker. Further, descriptively, the impact of Speaker gender appears greater for male Listeners in the data at hand.

Table 4.2 shows the mean number of LISTENER RESPONSES normalised relative to a 100-word turn by Listener and Speaker gender. For every turn, LISTENER RESPONSE frequency was calculated as [(number of LISTENER RESPONSES)/(number of words in turn)] x 100. 100 was used as a multiplier because it is close to the mean turn length of 103 words, but easier to interpret for the reader when analysing the

results. Thus, the numbers in table 4.2 can be understood as the mean number of LISTENER RESPONSES in a turn of mean length in this set of recordings for each Listener-Speaker gender combination.

Speaker Listener	Female	Male
Female	7.8 (sd = 18)	6.2
Male	5.8	4.5

TABLE 4.2
MEAN NUMBER OF LISTENER RESPONSES RELATIVE TO A TURN OF MEAN LENGTH BY SPEAKER AND LISTENER GENDER.

Table 4.2 shows that women produce approximately 8 LISTENER RESPONSES in an average-length Speaker turn when responding to another woman, while they only produce 6.2 in response to a male Speaker. Men listening to women produce on average 5.8 responses in a 100-word turn, a number fairly close to the women's response frequency in the mixed-gender interactions. However, when listening to a male Speaker, men's response frequency drops to an average of 4.5 responses. Recall that a note of caution is in order with respect to this last mean: while the amount of data for all-female and mixed-gender dyads is comparable, there is only one all-male dyad. Even bracketing this cell, the overview of means suggests cross-gender accommodation in terms of LISTENER RESPONSE frequency.

Additionally, figure 4.3 shows that there is large within-group variation for men Listening to women, and for women Listening to other women. The variation amongst women Listening to male participants is slightly smaller, and there is little variation in the all-male group. Given that there is only one all-male dyad this suggests a certain level of inter-speaker variation (or, to phrase it differently, that the frequency at which somebody produces vocalised LISTENER RESPONSES is somewhat idiosyncratic). Accordingly, we need to include Listener as a random effect in any statistical model on frequency.

To summarise and review the findings so far, the exploratory descriptive statistical analysis shows that the number of LISTENER RESPONSES increases with Speaker turn length. It also shows that Listeners only start responding once the ongoing turn exceeds a certain length. After that point, the relationship between turn length and number of LISTENER RESPONSES seems to be positive and linear. Furthermore, there seems to be cross-gender accommodation. However, given the small and unbalanced sample, this effect might be driven by only a few individuals, or confounded by turn length varying between groups. To shed light on what is driving this effect, the next section presents inferential statistics that take into
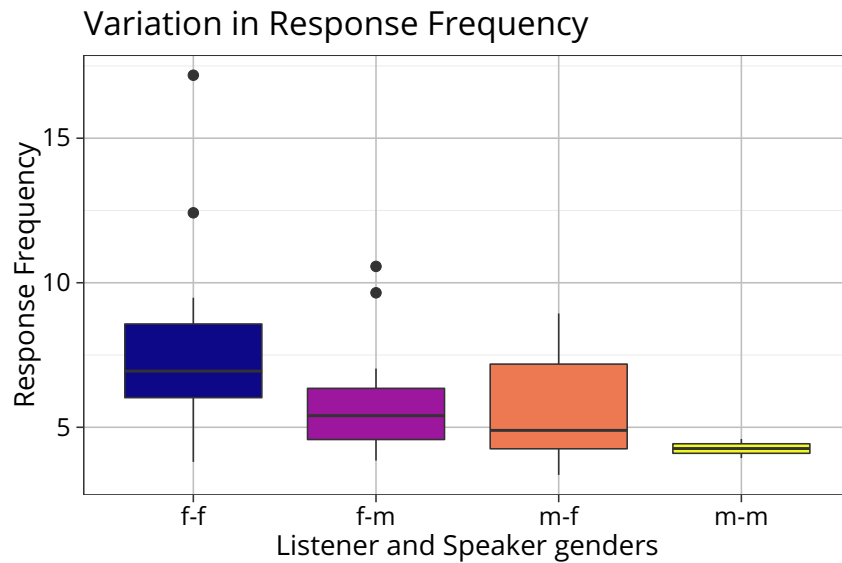
FIGURE 4.3
BOXPLOT SHOWING VARIATION IN THE RANGES AND STANDARD
DEVIATIONS OF LISTENER RESPONSE FREQUENCY BY LISTENER AND
SPEAKER GENDERS.

account the random effects of Listener and dyad as well as the fixed effects of turn
length and Listener and Speaker gender.

4.1.4   Inferential Statistics: Zero-inflated Poisson regression

Based on the definition of the variable and the exploration of the data presented
above, we have identified one linguistic and two social factors that influence
LISTENER RESPONSE frequency: the length of the ongoing Speaker-turn, and Listener
and Speaker gender. We have also noted the need for random effects for dyad
and Listener to account for within-group variability. In the following section I will
present a Zero-Inflated Poisson (ZIP) regression model which was fitted using the
GLMMTMB package (Brooks et al. 2017).

Zero-Inflated Poisson regression models are used to interpret count data with
a high number of zeros which could be related to different predictor variables. The
model consists of two parts: a logit (or zero inflation) model and a count model.
The logit model predicts *whether* we expect a LISTENER RESPONSE to be present by
estimating the number of 'excess zeros' in the data. The count model predicts *how
many* LISTENER RESPONSES we expect to observe.[4]

For the model, turn length is centred around the median of 75 words and
then log-normalised. Speaker gender is added as a fixed effect. The model also

---

[4]For other applications of poisson regression models in linguistics, see for example Coupé (2018),
Lo and Andrews (2015), and Winter and Wieling (2016).

takes into account that each individual behaves differently (random intercept for Listener) and that each dyad has its own dynamic (random intercept for dyad).[5] The same formula was used for both parts of the model. The formula is based on our qualitative and theoretical understanding of the data, rather than being built up or boiled down from a minimal or maximal model. The model summaries are given in table 4.3 for the logit model and table 4.4 for the count model. The model output is given in log units initially, but later converted to predicted counts in tables 4.5, 4.6, and figure 4.4 for ease of interpretation.

Fixed Effect	Estimate	Std. Error	Z-Value	p-value
(Intercept)	-8.14	1.64	-4.97	<0.001
words in turn (centred)	-1.12	0.05	-24.47	**<0.001**
Listener (male)	1.90	1.91	0.99	0.32
Speaker (male)	1.02	1.79	0.57	0.57
words in turn (centred) x Listener (male)	-0.23	0.08	-2.96	**0.003**

TABLE 4.3

FIXED EFFECTS OF THE LOGIT MODEL: THE LONGER THE TURN, THE MORE LIKELY THERE IS TO BE A LISTENER RESPONSE. THIS EFFECT IS SOMEWHAT LESS STRONG FOR MALE THAN FEMALE LISTENERS.

Fixed Effect	Estimate	Std. Error	Z-Value	p-value
(Intercept)	1.54	0.08	19.64	<0.001
words in turn (centred)	0.57	0.004	138.18	**<0.001**
Listener (male)	-0.07	0.11	-0.63	0.53
Speaker (male)	-0.16	0.08	-1.98	**0.048**
words in turn (centred) x Listener (male)	-0.01	0.01	-1.39	0.17

TABLE 4.4

FIXED EFFECTS OF THE COUNT MODEL: NUMBER OF LISTENER RESPONSES INCREASES WITH TURN LENGTH, AND MALE SPEAKERS RECEIVE FEWER RESPONSES THAN FEMALE ONES.

Both parts of the model take 'female' as the default gender and therefore report the effect of gender for the male Speakers and Listeners. The logit or zero inflation part of the model, presented first, predicts the occurrence of zeros, i.e. the likelihood of there *not* being a *Listener Response* at all reacting to any given turn. The logit model predicts that the likelihood of there being zero LISTENER RESPONSES decreases as turn length increases. The directions of the effects are as observed in figure 4.2: Male Listeners are more likely to produce no LISTENER RESPONSE at

---

[5]Ideally, we might also want to fit random slopes for each Listener and each Dyad. However, in a frequentist framework and with GLMMTMB just over 5200 LISTENER RESPONSES are not enough data points to do this.

all, and male Speakers are more likely to receive no LISTENER RESPONSE at all than women. However, neither of these effects is statistically significant. There is a significant interaction between turn length and Listener gender, though the effect size is relatively small. As turn length increases, male Listeners are less likely than female Listeners to produce no Response at all.

The random effects for the zero inflation model are Speaker (Variance = 1.603, sd = 1.266) and Dyad (Variance = 18.714, sd = 4.326), and for the conditional model Speaker (Variance = 0.016, sd = 0.128) and Dyad (Variance = 0.050, sd = 0.223).

The count model summarised in table 4.4 shows that turn length is a statistically significant predictor of number of LISTENER RESPONSES. As observed and logically expected, their number increases with turn length. Listener gender has no statistically significant effect, while the effect of Speaker gender is just above the 0.05 threshold. Thus, according to the count model, the observation that male Speakers receive fewer responses than female turn-holders is unlikely to be due to chance. However, considering that the p-value is just below 0.05 and the estimate is relatively small, this might be a Type 1 error.

We can use the model output to calculate how many words Speakers are predicted to utter before the Listener begins to respond (table 4.5).

Speaker Listener	Female	Male
Female	11.4	14
Male	12.2	14.8

TABLE 4.5
MODEL PREDICTION FOR THE LENGTH OF A TURN THAT RECEIVES 1 LISTENER RESPONSE: MALE LISTENERS WAIT FOR LONGER BEFORE THEY PRODUCE THE FIRST RESPONSE.

In the descriptive statistics it looked as though Listeners begin responding once the turns exceed approximately 10 words, with slight differences between male and female Listeners and Speakers. The model presented above predicts that Listeners begin to respond when the Speaker-turn has at least the number of words specified in table 4.5. Female Speakers are predicted to produce 11.4 words for a female Listener to produce one LISTENER RESPONSE (top left cell), and 12.2 words for a male Listener to respond (bottom left cell). Male speakers are predicted to say 14 words before a female Listener produces one response, and almost 15 words before a male Listener starts responding. We can also visualise the output of the count model to show *how many* LISTENER RESPONSES the model predicts based on turn length, Speaker gender, and Listener gender (figure 4.4).

The two statistically significant effects in these models are interesting with respect to previous findings about gender: (1) male Listeners wait longer until they

Predicted Listener Responses
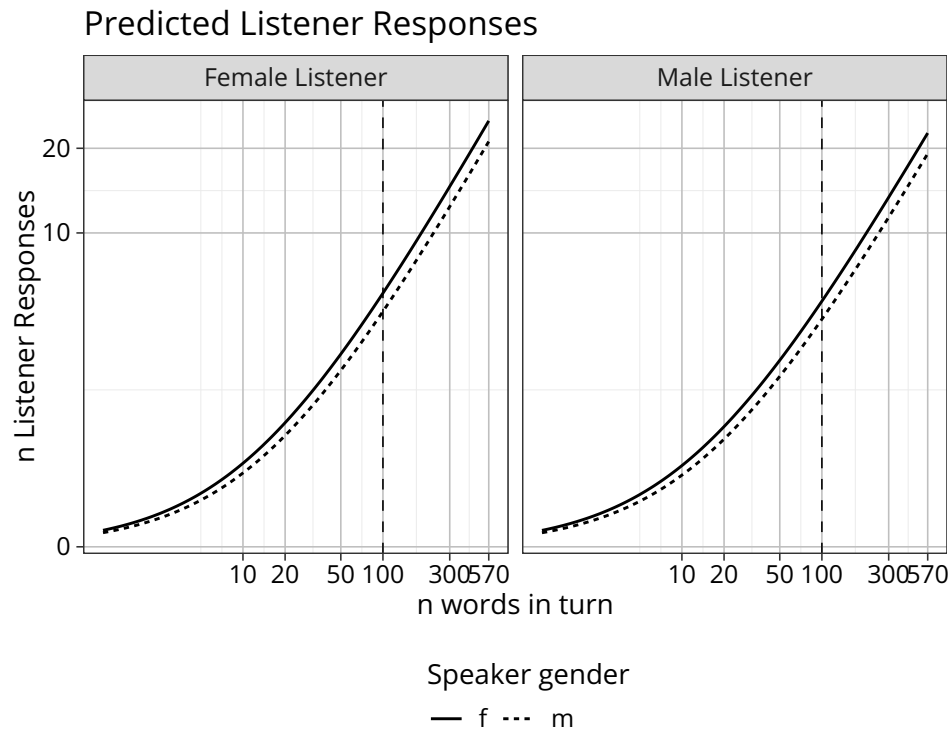
FIGURE 4.4

MODEL PREDICTION FOR THE NUMBER OF LISTENER RESPONSES BASED ON TURN LENGTH AND SPEAKER AND LISTENER GENDER. NUMBER OF RESPONSES INCREASES WITH TURN LENGTH, MORE SO FOR FEMALE THAN MALE LISTENERS AND SPEAKERS.

produce the first response, and (2) male Speakers receive fewer responses than female Speakers. (1) Could be seen to corroborate the notion of the 'unsupportive' male Listener, with men waiting longer until they do a LISTENER RESPONSE (see the argument made in Dixon and Foster (1998)). From a dominance perspective it could also be interpreted as men being more sensitive to giving up interactional power by ceding the floor, given that the first LISTENER RESPONSE can be considered a 'go-ahead' for the other person to produce an extended turn. Conversely, women could be seen as doing more social and interactional work by facilitating floor changes and producing more LISTENER RESPONSES than men.

Point (2) requires us to change our perspective from the Listener's to the Speaker's behaviour. Male Speakers are predicted to receive *fewer* responses than female Speakers, irrespective of Listener gender.[6] I have argued that in order for the Speaker to keep the floor, the Listener needs to signal their Listenership at

---

[6]I would like to note again that this effect is just above the significance threshold and should therefore not be over-interpreted. However, as we will see below, the model does take into account individual variation and predicts the group-level difference, so we also cannot disregard the finding.

more or less regular intervals. Accordingly, we can interpret the fact that male Speakers receive fewer responses than female Speakers as follows: male Speakers have a higher threshold for how much talk they can produce without receiving a response, and still legitimately keep the floor. Specifically, drawing on table 4.5, female Speakers 'need' a response after approximately 11-12 words, or the will cede the floor. Male Speakers continue speaking for 14-15 words until they 'need' a Listener Response. Thus, from a dominance perspective, this finding is in line with (1): Men are seen to behave in a way that reflects their (assumed) social and interactional power by keeping the floor for longer. They speak for longer without receiving the 'go ahead' to produce a multi-unit turn, and when they do hold the floor they wait for a longer time until ceding it to their interlocutor. However, from an interactional perspective these interpretations are premature; it is essential to investigate exactly *what* male and female Listeners do, and *how* they do it in the interaction to understand how potential power differences are performed and established in the interaction.

So let us now take a quick look at the within-group variation accounted for by the inferential model. If we compare the predicted number of Listener Responses to the observed, we get a first intuitive idea of just how great the impact of the individual Listener and dyad are. Table 4.6 summarises how many Listener Responses the model predicts relative to a 100-word Speaker turn, split by Speaker and Listener gender. In order to allow for a direct comparison, the observed values (see table 4.2) are given in brackets next to the predicted numbers. The top left cell of table 4.6 shows that the model predicts female Listeners to produce 6 responses in a 100-word long Speaker turn, while the observed mean was 7.8. The other cells are to be read in the same way.

Speaker / Listener	Female	Male
Female	6 (7.8)	5.1 (6.2)
Male	5.6 (5.8)	4.7 (4.5)

TABLE 4.6
MODEL PREDICTION (COMPARED TO OBSERVATIONAL DATA) FOR THE NUMBER OF LISTENER RESPONSES RELATIVE TO A TURN OF MEAN LENGTH (100 WORDS). THE HIGH FREQUENCIES OBSERVED IN FEMALE LISTENERS ARE DRIVEN BY INDIVIDUAL LISTENERS AND DYADS.

The comparison between the predicted and observed number of Listener Responses to a 100-word turn demonstrates two things: first of all, it concurs with the observation that women produce the highest number of Listener Responses when listening to other women, and men the lowest when listening to men. Both the observed and predicted values further suggest that there might be cross-

gender accommodation in terms of LISTENER RESPONSE frequency. Second, the predicted values are lower than the observed ones for all groups except the male-only dyad. Given that the ZIP-model includes random intercepts for Listener and dyad this suggests that the strength of the observed effect (particularly in all-female dyads) is driven by one or several individuals who produce LISTENER RESPONSES very frequently. Alternatively (or additionally), there might be one or several dyads in which the participants produce a very high number of LISTENER RESPONSES.

4.2   Discussion

The analysis presented here builds on previous work on variables above the level of the phoneme, particularly discourse-pragmatic ones, in defining the variable based on its position (see Waters (2016) and Ito and Tagliamonte (2003)). It also builds on Pichler's (2013b) proposal to integrate Conversation Analytic tools in the definition and analysis of discourse-pragmatic variables.

This chapter goes beyond these approaches and contributes to our conceptualisation of the variable, particularly the envelope of variation, in two ways: First of all, it shows that it is possible – and for some variables, like LISTENER RESPONSES, necessary – to go much further than using CA tools simply to get a better understanding of the local function of a variant; we can also base our approach to quantification on the interactional structure. Second, and in this vein, it shows that the envelope of variation is the multi-unit-turn produced by the Speaker rather than other talk done by the Listener, it offers a way of accountably quantifying frequency that does not require coding for absences.

I will briefly discuss both contributions in turn, mainly focussing on the question of quantifying based on interactional criteria. The investigation of LISTENER RESPONSES presented in this thesis begins at the highest level of abstraction, treating the variable as one cohesive category based on its sequential position as couched in and supporting ongoing Speaker-talk. Other approaches to defining variables above the level of the phoneme have proceeded from notions of functional equivalence[7] (Dines 1980; Lavandera 1978), or similarities in form or structure (see for example work on BE LIKE, LIKE, general extenders, or intensification), or the related concept of derivational equivalence (Pichler 2016b). Depending on the starting point, different aspects of the phenomenon can be construed as dependent and independent variables (see Waters (2016)).

---

[7]'Supporting the Speaker's ongoing talk' might be considered a kind of functional equivalence for LISTENER RESPONSES, but this is a very subjective way of defining the variable. Here, the focus is on the structural pattern which is intersubjectively reproducible with the help of the next-turn proof-procedure (see next paragraph).

The section on defining LISTENER RESPONSES as a variable has demonstrated that in this case a form-based definition includes lexical items that actually do a different type of interactional work and misses longer formulations that do the interactional work of supporting the Listener. The next-turn proof-procedure ensures that the coding is not based on researcher intuition but rooted in the participants' behaviour (see the extremely high inter coder reliability achieved, discussed in chapter 3). There has been much debate about how best to define the variable, but to my knowledge none of the scholars who have made a case for interaction-oriented analyses have extended this notion to the way they approach quantification.

This brings me to the second and more important contribution of this chapter. Given that LISTENER RESPONSES are rooted in the interaction, they need to be quantified in relation to their natural habitat: the turn they are responding to. This has implications for where we situate the 'linguistic system' in which the variation happens. In the past, this has unquestioningly been assumed to be the person producing the variable(s) under study, following Labov's (1972) seminal definition of the sociolinguistic variable. If we aim to more fully integrate an interactional perspective into our analysis of variation, we need to consider the possibility of the envelope of variation being not in the individual, but rather in the interactional space both participants co-create. This is nicely illustrated by an early – and easily resolved – miscommunication between Joe Fruehwald and myself: when revising my code and analysis, the 'mistake' Joe picked up on was that I was using Speaker-words as the envelope of variation, rather than the words produced by the person doing Listening. This exemplifies just how unusual and unintuitive it is for scholars trained in analysing language variation to situate the envelope of variation in the interlocutor or the interaction.

Listeners do not respond to their own talk, but rather to that of the Speaker. We see this reflected in Speaker gender having a greater – marginally significant – impact on the number of LISTENER RESPONSES in any given (multi-unit-)turn than Listener gender, which is not significant. Only one previous method is close to being interactionally accountable, albeit not theorised as such: Duncan and Fiske (1977) calculated number of LISTENER RESPONSES per 100 Speaker words (which, incidentally, is very close to the mean turn length of 103 words in my data). In their study, number of words and number of LISTENER RESPONSES were each totalled across the whole conversation, thereby obscuring possible effects of turn length. The approach presented here can be understood as an extension of their model. Interestingly, Duncan and Fiske's (1977) model has never been taken up in variationist studies of LISTENER RESPONSES. This might be at least partly due to the above mentioned orientation to situating the variable fully within one individual.

Past work on Listener Responses in different areas of linguistics notes the impact of the interactional context without presenting a clear solution. Specifically, Murphy (2012) analyses her data based on the Corpus Linguistic metric of words per minute and concludes that the gender differences she observes are actually due to an uneven distribution of Speaker and Listener roles in her data. Consequently, we cannot analyse variation based on social factors without somehow accounting for how Speaking and Listening are distributed in the interactions.

The three main approaches taken to quantifying Listener Responses in the past are (1) per million words in corpus, (2) per X minutes, or (3) relative to Speaker changes. Neither of these is particularly sensitive to the interaction and the distribution of Speaker and Listener roles. Furthermore, because they measure frequency in different ways, it becomes challenging to compare findings from different studies on the same social and linguistic variable. I will briefly present descriptive statistics for the three metrics and the operationalisation I have proposed (figure 4.5) and discuss how they compare.

In figure 4.5 the gender-composition of any given dyad is represented by the four groups on the x-axis. The scales of the y-axes vary between the four ways of quantifying because their scope and range are different.

Overall we can see that the approaches in the top three facets that are not sensitive to the interaction overstate how often male Listener respond to female Speakers, even for my fairly balanced data. The first facet in figure 4.5 shows the approach Murphy (2012) and other Corpus Linguists like O'Keeffe and Adolphs (2008) take. Murphy herself has concisely summarised the main issue with this approach: if one participant has the Speaker role most of the time and only very rarely does being a Listener, they will produce very few Listener Responses overall. If we quantify the number of Listener Responses based on all talk produced by *both* participants instead of the talk they are actually *responding to*, this confounds Listener Response frequency with role distribution.

The same reasoning applies to counting up the number of responses in a stretch of five minutes of interaction (see Maynard (1990) and Schweitzer and Lewandowski (2012)): if we do not know who is doing Speaking and who is doing Listening, the numbers might be skewed by an uneven distribution of roles.

The third approach, used for example in Clancy et al. (1996), Kogure (2003), and Brunner (1979), counts up the total number of Speaker changes and compares which percentage is to take over the floor, and which percentage is to do Listener Responses. The higher the Listener Response percentage, the more frequently that participant gives 'supportive feedback' compared to the number of times they take the floor. This metric certainly is more attuned to the interactional dynamics than the previous two, but it does not take the length of participants' individual

Comparing Normalisations
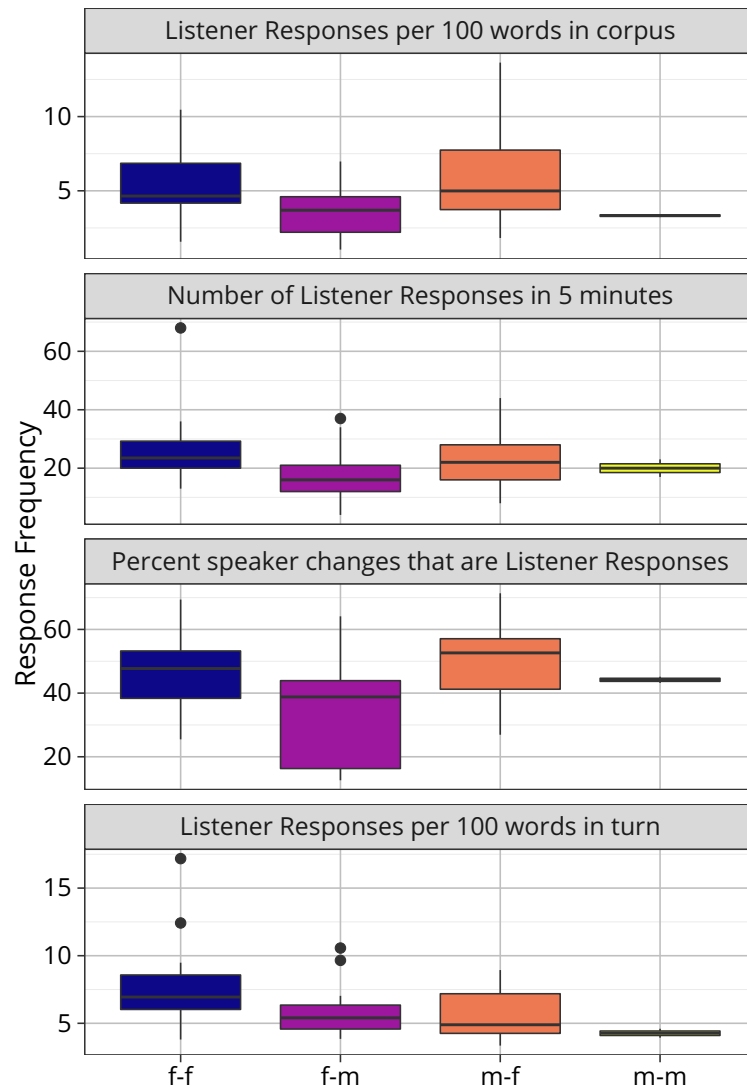
FIGURE 4.5

FACET PLOT COMPARING THREE WAYS OF OPERATIONALISING LISTENER RESPONSE FREQUENCY TO THE METHOD OUTLINED IN THIS CHAPTER. DIRECTION AND EFFECT SIZES VARY DEPENDING ON THE METHOD CHOSEN.

contributions into account. It could therefore still be skewed if one person was doing being a Speaker much more than doing being the Listener (for example in an interview situation).

Overall, as Murphy (2012) argues in her paper, metrics that are not interactionally sensitive produce skewed results, particularly when Listener and Speaker

roles are unevenly distributed.[8] As mentioned above, these different ways of operationalising frequency also have the practical implication that results from previous studies are not comparable.

To summarise, in this first analysis chapter I have defined LISTENER RESPONSES as a variable rooted in the interaction. Its definition is based on its position in the sequential structure, and the envelope of variation is the turn at talk that is being responded to rather than talk produced by the Listener. Acknowledging this and accounting for it in the quantification more fully integrates interactional and variationist analyses. Both of these theoretical and methodological contributions are only possible based on the next-turn proof-procedure and paying close attention to how the participants constantly (re-) negotiate their roles in the interaction.

The importance of the next-turn proof-procedure and its impact on the interpretation of our findings will become increasingly evident in the next chapters. The definition of LISTENER RESPONSES as a variable whose presence can be counted, while its absence cannot, is only the tip of the iceberg.

Interactional and CA scholars have critiqued past work for treating all LISTENER RESPONSES as the same (Schegloff 1982), because 'Listening' really is only a gloss for a number of different actions. Chapter 5 will therefore develop a taxonomy of LISTENER RESPONSES based on a close analysis of the sequential context preceding and following each LISTENER RESPONSE, and chapter 6 will show how this action type coding can serve as a basis for quantifying variation within LISTENER RESPONSES. Here, the definition of variable and variant shifts: The LISTENER RESPONSES that have been produced become the variable, and the individual action types are the possible variants.

Conservative CA scholars will object to any attempt at quantifying, siding with Schegloff (1993) in saying that we cannot account for all possibly relevant interactional details. However, I would like to argue that in cases where quantification is used to answer theoretically interesting and practically relevant questions, including even just a bit more interactional detail than before is a step towards a more accountable, locally relevant understanding of the variable in question. In this spirit, the approach to quantifying overall LISTENER RESPONSE frequency put forward in this chapter is a step towards a sociolinguistic theory that integrates interaction into our analysis of variation.

---

[8]In my corpus, these roles are fairly evenly distributed, and there are no institutional hierarchical difference or prescribed roles that would impact on the participants' conversations. If there were, the numeric differences between the four ways of operationalising frequency would be much greater.

Chapter 5

Not all Listener Responses are doing the same! A Taxonomy of Action Types

> There seems to be no agent more effective than another person in bringing a world for oneself alive, or, by a glance, a gesture, or a remark, shriveling up the reality in which one is lodged.

Erving Goffman (1972: 38)

One of the main points of critique of any quantitative work on Listener Responses from a Conversation Analytic perspective is that not all Listener Responses are created equal (going back to Schegloff (1982)). Indeed, Listeners can do a whole host of different *actions*,[1] and treating them all as the same obscures potentially interesting patterns of variation. CA scholars present this critique, but they do not present a viable way of addressing it. In fact, there is no cohesive Conversation Analytic overview of all the different actions Listeners can do and remain Listeners.

Instead, there are several disparate strands of literature: The foundational paper on *acknowledgement tokens* by Schegloff (1982) has sparked a debate about *acknowledgements* and Speakership incipiency based on the observation that some of the tokens used as Listener Responses can also be used to initiate a Speaker shift (Jefferson 1993; Drummond and Hopper 1993a). Another branch of work focuses on individual small tokens (or a handful thereof), and looks at the different things Listeners can do with them (particularly Gardner (1997), Gardner (1998), and Gardner (2001)).

---

[1]In this chapter, I will be referring to Listener Responses in small caps, in keeping with the notion that they are the variable, and to the individual *action types* in italics, to signal their status as variants – different ways of doing a Listener Response.

There is related work on small turn-initial particles and their various interactional functions, for example Sorjonen (2001) on *joo* and *nii(n)* in Finnish, and work on *oh* as a *marker of surprise* (Heritage 1984). At the other end of the spectrum, scholars have worked on individual actions which can sometimes be used as LISTENER RESPONSES – but this particular use is not foregrounded in these discussions. Examples are *assessment sequences* (Pomerantz 1984), *repair*, and *other-completion* (Lerner 2004b; Lerner 2004a).

If we think of *actions* Listeners can do as pieces of a puzzle, these are currently strewn around different tables, with no clear frame or guide on how to put them together. What I propose to do in this chapter is to use the definition of LISTENER RESPONSES based on their sequential position as a frame, and to put the jigsaw together so that we see the whole picture of LISTENER RESPONSE *actions*. A preliminary taxonomy was based on the work outlined in the literature review, and then further refined following an analysis of approximately 5200 vocalised LISTENER RESPONSES.

As outlined in the methodology and chapter 4, I propose a definition of LISTENER RESPONSES that is based on the turn-taking organization: A LISTENER RESPONSE is what the Listener can do in the Speaker's interactional space without disrupting their ongoing turn at talk, usually a multi-unit-turn. The taxonomy presented here contributes to Conversation Analytic work by showing that some actions not commonly considered LISTENER RESPONSES can be used as such, and how they differ from their use as full turns, and by pointing out that some LISTENER RESPONSES might have eluded more in-depth work so far. It further contributes to quantitative work on variation in LISTENER RESPONSES by offering a coding scheme based on categories locally relevant to the participants. This makes it possible to honour the diversity of LISTENER RESPONSE *actions* in a variationist analysis. The third analysis chapter will show an application of this coding scheme.

The following analysis will pay attention not only to the lexical form of the LISTENER RESPONSES, but also to their prosodic shape. Previous work has shown that LISTENER RESPONSES which align with the surrounding talk, and where the Listener-Speaker relationship is not challenged, generally prosodically match the host-TCU (Gorisch et al. 2012). We also know that prosodic design can cue interactional function, particularly for small, semantically void tokens (Barth-Weingarten 2011; Betz and Golato 2008). With respect to the LISTENER RESPONSES here, it can be observed that prosodic subordination is a strategy used to mark longer responses as LISTENER RESPONSES as opposed to attempts to take the floor. Further, Listeners can mark small tokens like *yeah* with pitch accents to cue that they do more than mere *acknowledgement* (indicating that the most frequent function of *yeah* in the sequential position of a Listener response is indeed acknowledging). Thanks to

their brevity, these responses cannot be mistaken as a claim for the floor unless followed by further talk, which means prosodic subordination is not needed to frame them as a Listening activity.

In the following analysis, I will first delve deeper into delimiting the phenomenon and then introduce the different actions Listeners can do based on a second longer extract. I will then elaborate on the analytic distinction between the different Action Types, drawing on more and shorter extracts. The examples chosen are representative of the participants' behaviours in the recorded interactions. It is common in Conversation Analysis to also discuss deviant cases to show that certain behaviours are oriented to as normative, and how deviations from the norm are resolved. Given that I do not present any new sequential structures or actions, but rather suggest a way of putting together a jigsaw whose individual pieces have already been well-described, I will focus on the representative. Discussions of deviant cases for the individual *Action Types* presented can be found in the literature cited. To conclude the chapter, I will present a tabular overview of the *Action Types* described. This table and the examples presented here are intended as a blueprint for a coding scheme to be used within quantitative studies of LISTENER RESPONSES.

5.1   Between Listening and Speaking

Let us start with an analysis of one longer extract, supplemented by examples from the extract just discussed in chapter 4. These initial data examples illustrate three main points: first, Listener and Speaker roles are constantly re-negotiated. Neither is taken for granted by the interactants, and both Listenership and Speakership need to be reaffirmed at more or less regular intervals. Second, the most common verbal feedback Listeners produce are *acknowledgements*. And third, there is a whole set of actions beyond acknowledgement that Listeners can do while reaffirming the current role distributions.

In the previous chapter, we have already encountered the overall structure of LISTENER RESPONSES:

STRUCTURAL DEFINITION OF A LISTENER RESPONSE

```
1    Speaker:    ongoing multi-unit-turn
2    Listener:   Listener response
3    Speaker:    continuation of multi-unit-turn
```

Here, I will first delve into delimiting the phenomenon in more detail than in chapter 4, focusing on cases of *incipient Speakership*, *third turn receipts*, and

distinguishing between LISTENER RESPONSES and full turns. I will do this mostly based on briefer examples from the long extract 1 presented in the previous chapter. I will then introduce the different *actions* Listeners can do based on some longer extracts, and finally delve into the individual Action Types in separate subsections for each.

5.1.1 Delimiting the phenomenon

The line between what is and is not a LISTENER RESPONSE is not always intuitive and clear, particularly if we are used to form-based or derivational descriptions of 'the variable'. I will therefore give three examples of *incipient speakership* and *third turn receipts* from the longer extract presented in chapter 4 to demonstrate that LISTENER RESPONSES really are defined based on interactional structure and not on their length or other aspects of form. LISTENER RESPONSES are everything Listeners do that supports the current role distribution and does not claim the floor. They are thus couched in the other Speaker's talk, though they might be comparatively 'long' or introduce a brief sideline that is needed to support the longer ongoing project (for example by requesting some additional information). I will first discuss an example of incipient speakership, then talk about third turn receipts, and then return to the overall question of delimiting LISTENER RESPONSES based on their sequential position.

Incipient Speakership

'Incipient speakership' is a term for the different signals participants employ to take over the floor or alert the current turn-holder to the fact that they would like to speak, with minimal disruption to the ongoing talk (Schegloff 1996). I will note examples of incipient speakership in the later extracts too, but here we will focus on the yeah in line 34 of the conversation between Donna and PuzzleB presented earlier.

> (2)   INCIPIENT SPEAKERSHIP, FROM EXTRACT 1, DONNA AND PUZZLEB, DIABETES CAMP
>   AND DAFNE, MINUTES 16:15-18:00

```
31        PuzzleB:      =it's just like 'Oh, being with other people who
32                      understand this and-'
33        PuzzleB:      (.5)
34  →     Donna:        yeah cause I'd never-
35        Donna:        well I'd met people-
36        Donna:        but I never knew anybody-
37        Donna:        that was diabetic until I did that DAFNE course,
```

```
38        PuzzleB:       /right oka[y\.
```

PuzzleB is moving towards a possible end of her narrative about her daughter's diabetes camp experience, and after a .5 second lapse (line 33), Donna takes the floor, starting her turn with the token yeah, which is often used as an acknowledgement. In this position though, it follows a pause and is immediately followed by more talk from Donna rather by talk from PuzzleB, who had the floor earlier. This makes it an instances of incipient speakership rather than a LISTENER RESPONSE – the latter would be couched in the Speaker's ongoing talk; it would elicit further same-Speaker talk rather than leading to the person who had the Listener role becoming the Speaker. I will point out more examples of this kind in the extracts discussed in this chapter.

Third turn receipts

Sometimes, what might be a LISTENER RESPONSE and elicit further talk is treated or used as a third turn receipt, specifically as an extension to a question-answer sequence, that then allows the person who asked the question to keep the floor. Usually, interactants produce so-called adjacency pairs, for example question-answer, invite-response, greeting-greeting. However, in some cases, the person who produced the first part of the adjacency pair, can produce a third part that extends the sequence. We frequently see this in class- or courtroom interactions, where the person in power asks a question (first part), receives a response (second part), and then receipts that response by for example confirming or paraphrasing it (third part and extension of the usual adjacency pair; see for example Antaki et al. (2000) and Lerner (1995), as well as Schegloff (2007) for a more general discussion of third turn extensions). This question-answer-receipt series is relatively common in some of the conversations and we will briefly look at an example where PuzzleB asks a question, which Donna answers, and then PuzzleB receipts the answer with a repeat of what Donna has said (lines 48-50). This repeat stands in overlap with Donna's response, and following it, PuzzleB immediately launches a new turn of her own, prefaced with oh right (line 51).

(3)   THIRD TURN RECEIPT AND INCIPIENT SPEAKERSHIP, FROM EXTRACT 1, DONNA AND
      PUZZLEB, DIABETES CAMP AND DAFNE, MINUTES 16:15-18:00

```
48        PuzzleB:       and how long ago did you go on that course,
49        Donna:         [!] um:: two years ag[o-
50  →     PuzzleB:                             [two years ago.
51        PuzzleB:       ↑oh right so that's quite good isn't it that there's
52  →                    still the:-
```

```
53      Donna:      yeah:[:-
54      PuzzleB:         [yeah and the-
55      PuzzleB:    yeah cause that's what happened my daughter after going
56                  on the: um camp is that-
```

These Question – Answer – Receipt formats are well attested and very common in situations like for example (news) interviews and classroom interactions (see Schegloff (2007) and Lerner (1995)). Importantly here, the receipt and the `oh right` following it, which based on their form might have been considered potential LISTENER RESPONSES are not considered part of this set of actions because they do not support the Speaker-Listener role-distribution that might have been initiated by PuzzleB's question in line 48 – this could have launched a longer telling from Donna, but instead PuzzleB keeps the floor after Donna's brief response.

I have briefly presented three examples of cases where parts of turns or individual contributions by the person who has been the Listener for a while potentially look like LISTENER RESPONSES at the level of form, but in fact do something else – either indicate that the person who has been Listening wants to take over the floor, or simply receipt the answer to a question, before the person who has asked the question continues (as in the extract just discussed). The example just discussed nicely contrasts with lines 11 and 13/14 in the conversation between Samantha and Lavina (extract 5) which I will present to introduce the diversity of actions Listeners can do.

Sequential function > length

We have seen that brevity does not a LISTENER RESPONSES make – it is their sequential positioning and impact, couched between and supporting the turn-holder's ongoing talk. Similarly, utterances are not disqualified from the set of LISTENER RESPONSES simply because they are not 'brief enough'. If LISTENER RESPONSES are defined based on the sequential structure and the next-turn proof-procedure, their status does not hinge on their length. Examples of different lengths will be discussed throughout this chapter, particularly in the section on *joint constructions*.

I will present a slightly longer extract with an instance of *voicing* here. Tess is telling Lily about a consultant she used to have as a teenager, and how he responded to her eating cake on her brother's birthday:

(4) EXAMPLE OF LONG LISTENER RESPONSE (VOICING) FROM TESS AND LILY, MIN 12:00-13:00

```
1       Tess:       [or I ] used to have a (.) I used to have a consultant-
2       Tess:       and he's one of these guys who (.) he just like-
```

111

```
3      Tess:          d'you know, was a health freak of er in of his own
4                     accord=
5      Tess:          =[he wasn't diabet]ic or anything [but he us]ed to-
6      Lily:           [mhm-          ]                 [mhm-    ]
7      Tess:          he'd get up like half six every morning and go on like
8      Tess:          a twenty kay bike [ride before coming to work] and-
9      Lily:                            [wow-                       ]
10     Tess:          only ate salad and never had any (.)
11     Tess:          like dessert or treats or anything like that.
12     Tess:          and then I'd go in-
13     Tess:          (.6)
14     Tess:          [!] and I'd have my diary and he was saying
15     Tess:          god you (.) went up after your dinner this day-
16     Tess:          [or whatever and like what's happened-]
17     Lily:          [((inhale - exhale))                 ]
18     Tess:          I'd say was my brother's birthday like we had cake.
19     Tess:          (.) and did you eat some of the cake?
20     Lily:          [((laugh))]
21     Tess:          [yes.     ]
22     Tess:                        [why?      ]
23 →   Lily:                        [no (.) It] sat on
24 →   Lily:          [my plate and I looked at it.                    ]
25     Tess:          [((really)), yeah it sat there and there was like=]
26     Tess:          =by diffusion I became high [after it no.]
27     Lily:                                      [((laugh))   ]
28     Tess:          um-
29     Tess:          (.) yes I had some.
30     Tess:          why (.) why would you do that,
31     Tess:          'n I'm like It's my brother's birthday=
32     Tess:          =and there was birthday cake-
33     Lily:          ((laugh))
34     Tess:          (.5)
35     Tess:          but you know that it's going to affect your blood sugar
36                    and I was like-
37     Tess:          yeah, and it didn't KILL me [like I had one=]
38     Lily:                                      [((laugh))      ]
39     Tess:          =slice of cake and he literally was just like-
40     Tess:          cannot compute do you know what I [mean,]
```

Tess is narrating this story about her consultant – she produces the vast majority of the talk, with Lily providing different types of Listener Responses throughout, most of them very brief. They include short vocalisations like `mhm` and `wow` in lines 6 and 9, as well as laughter. Towards the climax of the narrative, when Tess quotes the conversation between the consultant and herself, Lily voices a possible response Tess might have given her consultant (lines 23 and 24, marked with arrows). Most of the voicing overlaps with Tess' ongoing talk, and Lily makes no attempt at taking over the floor; once she has uttered the potential one-sentence-answer she signals her continued Listenership with laughter, represented in lines 27, 33, and 38.

This type of voicing in the structural sequence we see here is considered a Listener Response because it makes no attempt at taking over the floor – it occurs in overlap with the ongoing talk, is prosodically matched to it, and followed by more prototypical brief Listener Responses that reaffirm the Speaker-Listener role distribution. If the first two sections on delimiting the phenomenon have shown that a contribution is not a Listener Response just because it is brief or has a certain lexical shape, this section has illustrated that a contribution can not be excluded from the set of Listener Responses only because is comparatively long. The definition is based on sequential structure, not on form. This structural definition of Listener Responses has been visually introduced in 2.1 in the Literature Review, and concisely summarises the theoretical and methodological underpinnings of the definition and delimitation of the variable as discussed here, and the variants (ie the individual Action Types), which we will see in the next sections.

5.1.2   Delving into the diversity of actions

As we have already noted in the first long extract, Listeners can do a number of different *actions*. *Acknowledgements* are by far the most frequent, and we have encountered them for example on lines 38 and 53 in extract 1. We have also encountered the participants *jointly constructing a turn* (line 10) without a floor-change, the Listener doing a *surprise mark* (line 45), and proffering a *first assessment* (line 30), both discussed in chapter 4.

In order to make it easier to identify the different action types at a glance, they are colour-coded in the following longer extract. Later, arrows will be used to indicate the lines under discussion. The colour scheme is as follows:

- Acknowledgements

- incipient speakership – not a Listener Response

- Surprise Mark

- First Assessment

- Second Assessment[2]

- Self-initiated other-repair

- Other-initiation

- Voicing

- Collaborative Completion

This second longer extract also contains a large number of *acknowledgements*, which I shall not comment on in detail here. The following section will focus on those Listener actions that we have not seen in the first extract. They are both colour-coded and marked with arrows to make them easier to spot because this is a fairly long extract. It stems from early on in the conversation about healthcare and technology between Samantha and Lavina. Samantha prompts a multi-unit-turn from Lavina by asking how technologically advanced her treatment is, and Lavina responds by discussing not only her devices (a specific CGM, the Dexcom, and an insulin pump) but also the DAFNE course she participated in.

(5) LONG EXTRACT 5, LAVINA AND SAMANTHA, TECHNOLOGY AND DAFNE, MINUTES 02:30-04:00

```
1        Samantha:      so how technologically advanced is your treatment,
2        Lavina:        [!]
3        Lavina:        (.4)
4        Lavina:        I:: self-fund.
5        Lavina:        so I ha::ve the dexcom.
6                       (.5)
7        Samantha:      °right°.
8        Lavina:        [!] um:-
9        Lavina:        (.7)
10       Lavina:        I:: started [using-
11  →    Samantha:                  [you're a pump user.
```

---

[2]Surprise marks, first, and second assessment are listed here for the sake of completeness but do not appear in the longer extract presented next. We saw some instances in the first long extract in chapter 4 and will see shorter examples in the relevant subsections in the present chapter.

```
12        Lavina:      yes.
13        Samantha:    yes.
14        Samantha:    °°mhm°°.
15        Lavina:      but I wasn't (.) whe:n I got the libre system.
16        Samantha:    °uh-huh°.
17        Lavina:      u:m:-
18        Lavina:      (1.2)
19        Lavina:      so:: then-
20        Lavina:      (.8)
21        Lavina:      I:: did a DAFNE-
22        Lavina:      the DAFNE [training course in ↑f::ebruary twenty ten-
23        Samantha:             [/yes-
24        Samantha:    °°mhm°°.
25        Lavina:      a::nd (.) that highlighted what I'd been-
26        Lavina:      telling my diabetes team for twenty odd years.
27        Lavina:      that there was-
28        Lavina:      a problem being created-
29        Lavina:      by the regime that I was on.
30        Samantha:    °right°.
31        Lavina:      um::-
32        Lavina:      (.6)
33        Lavina:      but it wasn't until I did DAFNE-
34        Lavina:      (.7)
35        Lavina:      that straight away-
36  →     Samantha:    [it highlighted it.
37        Lavina:      [when my readings were put [on the board-
38        Samantha:                                [°°mhm°°.
39        Lavina:      the nurses that I'd been seeing for twenty-odd years
40                     said-
41        Lavina:      'yes, we think you're having-
42        Lavina:      ↑nighttime [hypos.'
43        Samantha:               [°yes°.
44        Samantha:    °mhm°.
45        Lavina:      (.6)
46        Lavina:      'cán you gét yourself úp at twó o'clock in the
47                     mórning-'
48  →     Samantha:    'nó I cá:n't (.) thán[k you [í've got a jób to gó to.'
49        Lavina:                           [!]   [no I'd-
50        Samantha:    [((laughing))
```

115

```
51        Lavina:     [I- did do tha[tʰ-
52        Samantha:                  [°°mhm°°-
53        Lavina:     um::-
54        Lavina:     so starting on the first of february twenty-ten-
55        Lavina:     (.8)
56        Lavina:     I:::-
57        Lavina:     (.5)
58        Lavina:     then (.) didn't-
59        Lavina:     (.5)
60        Lavina:     sleep-
61        Lavina:     (.8)
62        Lavina:     fully-
63        Lavina:     (.6)
64        Lavina:     [any night of the year.
65        Samantha:   [you always woke at twelve and m-.
66                    (.4)
67        Samantha:   a[t two-
68        Lavina:      [so:-
69        Lavina:     whe:n the libre system came out (.) that wa:s-
70        Lavina:     fanta:stic for me-
71        Lavina:     cause I could literally just scan myself in the=
72        Lavina:     =[morning-
73        Samantha:    [°°yes°°-
74        Lavina:     and getʰː-
75        Lavina:     (.4)
76  →     Samantha:   [°°a pattern°°-
77        Lavina:     [a rough idea [of what the trend was.
78        Samantha:                 [°°mm°°-
79        Samantha:   °°mm hm°°-
80        Lavina:     not so much the numbers but the trend.
81        Samantha:   °°yeah°°-
82        Lavina:     um::-
                      30 SECONDS OMITTED
83        Lavina:     and at that stage I also had the joy of a marked dawn
84                    phe↑nomenon,
85  →     Samantha:   °a wha-°/yes-
86        Lavina:     so I:: would be up in-
87        Lavina:     twenty-two twenty-[three in the morning,
88        Samantha:                     [but that's also the compensation=
```

116

```
89      Samantha:       =for the hypo in the [middle of the night.
90      Lavina:                               [yes.
91      Lavina:         ex[actly.
92      Samantha:          [°mhm°.
```

LAVINA CONTINUES

The LISTENER RESPONSE *actions* I will discuss with respect to this extract are *other-initiation* (line 11), *self-initiated other-repair* (line 36 and 76), *voicing* as seen in the previous extract (line 48), and *other-initiation* as opposed to *acknowledgement* (line 85).

The sequence is launched by Samantha asking Lavina about her technological involvement, and Lavina begins her answer with I have the dexcom (line 5), not commenting on insulin pumps at all. Following several hesitations from Lavina's side, Samantha then *initiates repair* with the request for information you're a pump user (line 11). This *repair-initiation* problematises Lavina's response to the initial question as incomplete and lacking one specific detail. It is produced in overlap with Lavina's TCU I started using (line 10), at a point where the TCU could be continued either with reference to the CGM or with respect to other diabetes technologies. Samantha's *repair-initiation* is prosodically not reduced, but it is brief and Samantha follows it up with acknowledgements (lines 13, 14) as soon as Lavina launches her answer, thereby reaffirming her position as a Listener. Having provided the requested information, Lavina goes back to her starting point that she starting using the Freestyle Libre before she was given an insulin pump. This makes the *repair initiation* in line 11 (which is a request for information) different from the question-answer sequences I have presented above as *not* Listener Responses: Samantha's request for information here leads to Lavina providing the information necessary for Samantha to remain a 'good' Listener, and for Samantha's mind being able to 'stay with' Lavina's talk. If Samantha had followed it up with another question, Lavina provided another answer, and so on, then the distribution of the roles and the floor would have changed. However, we can see that both participants treat the *repair initiation* in line 11 as an activity that supports Lavina's ongoing telling, which continues for another 80 lines here.

In lines 15-29 she describes how the analysis of her glucose readings during the DAFNE course made it evident that her regime of injections was not working. She expands on this from lines 31 onwards, but her speech becomes less fast and she leaves longer pauses between the increments in lines 31, 33, and 35. At the end of that third increment which is syntactically and prosodically incomplete, Samantha provides a syntactically matched possible continuation of Lavina's turn

in line 36. This could be interpreted as another type of repair, called *self-initiated other-repair*: Lavina's repeated hesitations indicate trouble formulating the next TCUs, and Samantha offers a candidate continuation. Lavina however continues in overlap with Samantha, presenting a slightly different formulation. Samantha produces an *acknowledgement* (line 38) as soon as she has finished her possible continuation. Lavina does not comment on Samantha's suggestion for completion, but her following talk shows that Samantha's formulation captured the essence of the story by reporting how the nurses reacted to seeing Lavina's glucose readings on the board.

Lavina moves the story forward to how the nurses suggested she get up at two o'clock every night to test her blood sugar. Here (line 48), Samantha *voices* an answer Lavina might have given to her team – the same Listener response action we have seen in the extract 1, where Donna provided a possible continuation of PuzzleB's response to her daughter. The rhythmic design of Samantha's contribution closely matches Lavina's preceding turn. Rhythm is indicated with an accent over each accented syllable. Note in particular the contrastive stress between the 'can you' in Lavina's turn, and the 'can't' in Samantha's candidate completion. The second turn is also prosodically matched in the sense that neither of the contributions has strong pitch excursions. In line 49, Lavina produces a click after `no I can't` in overlap with `thank`, starts to say `no I'd` at the next TRP in overlap with `I've got` and then says `I did do that`^h (line 51) in overlap with Samantha's post-completion laughter (line 50), effectively rejecting Samantha's suggestion.

Samantha moves from laughing (line 50) to `mhm` (line 52) in overlap with the release and aspiration of the stop at the end of Lavina's TCU and Lavina continues her DAFNE and diabetes technology story. These two longer LISTENER RESPONSES on lines 36 and 48 from the conversation between Lavina and Samantha show a clear orientation to Lavina being the main Speaker and holding the floor. Lavina makes contributions at all possible TRPs in this second case (despite Samantha not having completed her Listener turn yet), and Samantha immediately reconfirms her Listenership with an *acknowledgement* when she has finished her possible continuation. Lavina then continues the narrative leading up to how the Freestyle Libre changed her life because instead of having to wake up several times at night she could simply scan the sensor in the morning and get a complete log (line 71).

Again, Lavina has been delivering her talk fluently, with no interruptions at all. She hesitates at `and get`^h: – (line 74), leaving a .4 second pause (line 75) until Samantha comes in and suggests `a pattern` (line 76) as a possible completion. There are no other trouble indicators in Lavinas preceding talk, but given the absence of pauses or other hesitations in her previous turns, this pause can be

taken as a clear hesitation, particularly at a point of projectable incompletion. The strong release and aspiration of the voiceless plosive at the end of get^h: further indicates this is a 'holding pause' (Local and Kelly 1986). Samanthas word supply is prosodically reduced relative to Lavinas talk, its amplitude is low and there is no pitch movement on it. Samantha immediately produces a continuer mm (line 78) when Lavina continues in overlap with her word supply, and another one at the end of Lavinas continuing TCU (line 79), clearly signalling their Speaker-Listener relationship still holds.

Towards the very end of this long extract, Lavina states that at the time she also had a marked Dawn Phenomenon (line 82,83). This is a technical term, and Samantha begins to initiate repair with a wha- in line 84 but then self-corrects, producing an *acknowledgement* instead. We will return to this example in the section on *other-initiation*.

To summarise, we have seen the following Listener Responses in the two long extracts: First, *acknowledgements*, which are the most common, ubiquitous, and unmarked Listener Responses. Second, *assessments* – both first and second – which we are familiar with from work by Pomerantz (1984). These action types (as well as markers of surprise, which we will see examples of soon) have been discussed in previous work on Listener Responses, for example by Schegloff (1982), Jefferson (1984a), and Gardner (2001). Listeners can further initiate or *do repair*; so far only *repair-initiation* has been described as a Listener activity (Gardner 2001). And finally, Listeners can voice a character in the main Speaker's story or complete the main Speaker's sentence (see Lerner (2004b)); both are actions that have not received much attention as Listener activities.

Previous approaches have looked at a specific set of lexical items and analysed all their functions in different sequential positions – for example all uses of *yeah* or *oh*. A subset of things participants can do with these tokens are Listener Responses, but papers focusing on individual tokens will not see the whole picture of Listener actions. Even work focusing on Listener Responses has generally cast its net based on form and thus excluded longer contributions like the examples of repair or voicing given above. I hope to have shown that all the actions discussed above follow a common structure as Listener Responses that makes them part of a cohesive set of actions.

I have already gestured at the fact that these Listener response *actions* differ with respect to the TCU they respond to, and the kind of continuation they make relevant. In the next sections, I will present analytic distinctions between the categories we have seen in extracts 1 and 5.

5.2 Acknowledgements

Acknowledgements are the most prototypical and most extensively described LIS-TENER RESPONSES, and they make up about 80% of all LISTENER RESPONSES in the present corpus. They overlap with the ongoing Speaker-turn more often than other LISTENER RESPONSES, frequently following the end of one TCU in a multi-unit-turn and overlapping with the beginning of the next. They can also stand in the clear between two TCUs. They are both backward- and forward-looking in that they acknowledge the preceding talk and orient to the relevance of the current Speaker continuing without making any particular kind of continuation relevant. In terms of turn-taking structure, this can be summarised as follows:

STRUCTURE OF ACKNOWLEDGEMENTS AS LISTENER RESPONSES

```
1    Speaker:    ongoing multi-unit-turn
2    Listener:   acknowledgement
3    Speaker:    continuation of ongoing multi-unit-turn
```

This pattern has been well described in the previous literature, and my data follow what has been attested. From a structural point of view, it becomes obvious that these LISTENER RESPONSES orient to the distribution of Speaker and Listener roles, not to the specific content of the talk. They are what Bavelas et al. (2002) class as 'generic backchannels': they respond to unremarkable talk (or treat the talk they respond to as unremarkable), simply signalling continued Listenership without taking any particular stance to the ongoing turn. We have seen this for example in lines 20-22 in extract 1:

(6) ACKNOWLEDGEMENT, FROM EXTRACT 1, DONNA AND PUZZLEB, DIABETES CAMP AND DAFNE, MINUTES 16:15-18:00

```
20        PuzzleB:       but I think a lot of them'd been sent-
21  →     Donna:         °mhm°.
22        PuzzleB:       because they weren't controlling- (.)
```

In these conversations, laughter is treated exactly like more lexically realised versions of acknowledgement, as we have seen in the first lines of extract 1. For a discussion of laughter as a LISTENER RESPONSE see for example Namba (2011).[3]

Lines 41-46 from extract 5 demonstrate that *acknowledgements* make further talk relevant: Samantha's first *acknowledgement* in line 43 overlaps with the end

---

[3]Subtleties in the timing and exact design of the laughter tokens merit their own in-depth discussion, in fact, Namba (2011) is a whole PhD thesis focused on this specific question. Because there are only 50 instances of laughter in the 5200 responses analysed here, I treat them as one group and leave their separate analysis for future work.

of Lavina's talk, she then produces a second *acknowledgement* (line 44), and does not make any move to take the floor in the ensuing .6 second lapse before Lavina continues her narrative:

(7) ACKNOWLEDGEMENTS, FROM EXTRACT 5, LAVINA AND SAMANTHA, TECHNOLOGY AND DAFNE, MINUTES 02:30-04:00

```
42       Lavina:         ↑'nighttime [hypos.'
43  →    Samantha:                   [°yes°.
44  →    Samantha:        °mhm°.
45       Lavina:         (.6)
46       Lavina:         '/can you get yourself up at two o'clock in the
47                       morning?'
```

As Schegloff (1982) points out, *acknowledgements* pass an opportunity to initiate repair and treat the previous talk as unproblematic. This is exactly what we can see at the end of extract 5 where Samantha begins to *initiate repair* but then self-corrects and produces an *acknowledgement* instead, inviting Lavina to continue without needing to address any trouble source.

(8) ACKNOWLEDGEMENTS, FROM EXTRACT 5, LAVINA AND SAMANTHA, TECHNOLOGY AND DAFNE, MINUTES 02:30-04:00

```
83       Lavina:         and at that stage I also had the joy of a marked dawn
84                       phe↑nomenon,
85  →    Samantha:        °a wha-°/yes-
86       Lavina:         so I:: would be up in-
```

The next brief example demonstrates that sometimes preceding speech makes relevant more than *acknowledgement*, and Speakers will pursue a specific kind of LISTENER RESPONSE if only an *acknowledgement* is offered. It is taken from early on in the conversation between Emma and Lavina.

(9) WHEN ACKNOWLEDGEMENT IS NOT ENOUGH, EMMA AND LAVINA, PRIORITISING DIABETES MANAGEMENT, MINUTES 01:10-01:30

```
1        Lavina:         I: now would say that I'm in a very unusual position=
2        Lavina:         =in that I can prioritise my diabetes-
3        Lavina:         (1.0)
4        Lavina:         a::nd in some wa:ys-
5        Lavina:         that's been a bit of an eye-opener for me.
6   →    Emma:           °mm°-
7   →    Emma:           (.4)
```

```
 8  →    Emma:         I bet.
 9       Lavina:       [!]
10       Lavina:       (.5)
11       Lavina:       u::m-
```

LAVINA CONTINUES

Lavina states that having been able to focus solely on her diabetes after her early retirement has been an eye-opener for her (lines 1-5). Emma responds to this with the minimal *acknowledgement* `mm` (line 6). Usually, *acknowledgements* like this one are almost immediately followed by more Speaker-talk, but here a pause of .4 seconds unfolds (line 7), before Emma proffers an epistemically downgraded *agreement* (`I bet`, line 8) with Lavina's preceding statement. At this point Lavina initiates her continuation with a dental click, but there is another unfilled pause and a filled pause `u::m` before she does launch the continuation of the ongoing project. This example shows that sometimes more than mere *acknowledgement* is required from Listeners; they need to take a stance or make an *assessment*. We will see more complex examples in later sections.

To summarise, *acknowledgements* respond to preceding talk as unproblematic to hear, understand, and receive, as unsurprising and not making a more involved reaction relevant. *Surprise marks*, in contrast, treat something in the preceding talk as new, surprising, or otherwise remarkable. They usually elicit more on-topic talk elaborating on the aspect that has been marked as surprising (see also Tolins and Fox Tree (2014)).

5.3 Surprise marks

As just demonstrated, sometimes more than just *acknowledgement* is made relevant by the Speaker-talk. *Surprise marks* are one of those response actions. They are similar to *acknowledgements* in their brevity, but different in terms of interactional context and prosodic marking. *Surprise marks* are signals Listeners use to show their appreciation of the content of a previous TCU as new, surprising, or otherwise unexpected. They are most often placed at the very end of the TCU and frequently overlap with the beginning of the new TCU in which the main Speaker elaborates on the relevance or impact of the news. In terms of their sequential structure we can characterise them as follows:

STRUCTURE OF SURPRISE MARKS AS LISTENER RESPONSES

122

```
1    Speaker:    ongoing turn (containing something surprising, unusual, or
2                unexpected)
3    Listener:   surprise mark
4    Speaker:    continuation of ongoing turn (often elaboration of the thing
5                marked as surprising)
```

The following examples will show how *surprise marks* as a Listener response differ from *acknowledgements*. The first example is taken from the conversation between PuzzleB and Donna. PuzzleB has just told Donna how her daughter's teachers problematised the girl eating crisps and other junk food at a class trip. PuzzleB argues her daughter can eat everything she wants as long as she takes the adequate amount of insulin for it. This is where Donna takes over the floor to give an example from her childhood that frames this dietary freedom as relatively new.

(10) Markers of Surprise, Donna and PuzzleB, Birthday Parties on MDI, minutes 10:40-11:00

```
1         PuzzleB:      you [know(.) but but you just-
2         Donna:            [I used to find that hard when I was young- li-
3                       like-
4         Donna:        children's parties when I was first diagno:sed-
5         Donna:        be[cause it wasn't carb counting.
6         PuzzleB:        [°right°.
7         PuzzleB:      [°yeah°,
8         Donna:        [I kinda had to eat the same amount every day,
9  →      PuzzleB:      /oh::[:\ right °okay°.
10        Donna:             [it was more cause it was a set dose-
11        Donna:        you were meant to eat roughly [the same amount er-
12        PuzzleB:                                    [°right okay°.
13        Donna:        (.6)
14        Donna:        and so at parties and stuff-

                        DONNA CONTINUES
```

PuzzleB's Listener response oh right okay immediately follows Donna's revelation that as a child she had to eat the same thing every day. The oh is in the clear, but the rest of the response overlaps with Donna's continuation. The Listener response is prosodically downgraded relative to the main Speaker talk, it is quieter and there is little pitch movement, clearly marking it as not Speakership incipient. There is a pitch rise-fall on the oh, marking it as the central element of this response. Donna does not make any explicit comment on the Listener

```

RESPONSE and simply continues to rephrase what she has just presented in the TCU that PuzzleB marked as surprising.

PuzzleB then responds to this elaboration with an *acknowledgement* `right okay`, which is lexically very close to the *surprise mark* but differs from it in two central respects: it is not preceded by the prototypical token *oh* (Heritage 1998), and there is almost no pitch movement, while there was a pitch accent on `oh` in the *surprise mark*. Furthermore, the *acknowledgement* is uttered during an ongoing TCU, while the *surprise mark* was produced after TCU completion. In response to the *acknowledgement*, Donna moves her story forward, while the TCU following the *surprise mark* elaborated on the TCU marked as surprising.

(11)   MARKERS OF SURPRISE, TESS AND VELOMINATI, URINE ANALYSIS, MINUTES 2:55-3:10

```
1   Velominati: ba:ck in the da:y-
2   Tess: °°((laugh))°°
3   Velominati: um::-
4   Velominati: (.5)
5   Velominati: [!] the- there was- there were no glucose monitors-
6   Velominati: ((that that))-
7 →   Tess: [/wow\.
8   Velominati: [so it wasn't an option (.) wasn't available.
9   Tess: °o/kay°.
```

VELOMINATI CONTINUES

The next example is taken from the conversation between Tess and Velominati, in which Velominati tells Tess how he dealt with his diabetes during his time at university right after diagnosis. Tess' *surprise mark* is a lone-standing `wow` that follows Velominati's TCU and overlaps with his continuation, which makes explicit the consequence of the fact that there were no glucose monitors. The `wow` has a lower amplitude than Velominati's talk and it carries a slight prosodic rise-fall, but no strong prosodic marking. It is minimal and Tess makes no claim to the floor. Tess then produces an *acknowledgement* in response to Velominati's continuation.

Tess' `wow` might be mistaken for an *assessment*, but it is prosodically less marked than a minimal *first assessment* in LISTENER RESPONSE position – such a minimal *first assessment* would carry a stronger prosodic marking. Alternatively a prosodically less marked *first assessment* would be more elaborate, following the structural formula outlined in the next section.

The next example also shows that *surprise marks* are not *first assessments* – here the *marker of surprise* is followed by a *first assessment*. The extract stems

from the conversation between Donna and KinderSurprise. KinderSurprise shows Donna her diabetes technology; a Freestyle Libre she is currently trialling, and the insulin pump that she has had for a few years now. Both are new to Donna.

(12)  MARKERS OF SURPRISE, DONNA AND KINDERSURPRISE, PUMP AND METRE LINKING, MINUTES 15:30-15:40

```
 1          KinderS:    I quite like that cause that links up to my pump,
 2  →       Donna:      [!] ↑oh::-
 3          KinderS:    [so if I just took my sugars it sends it by-
 4          KinderS:       [bluetooth and tells my pump (.) what my sugars are,
 5  →       Donna:          [°that's good°.
 6          KinderS:    so it's quite-
 7          KinderS:    (.4)
 8          KinderS:    that one doesn't-
 9          KinderS:    (.7)
10          KinderS:    do that.
```

Donna's *surprise mark*, an oh preceded immediately by a dental click, follows right after KinderSurprise informing her that the glucose metre links to the insulin pump. It overlaps with KinderSurprise elaborating further on how exactly this works. Donna responds to these details with a *first assessment* that's good, also in overlap with KinderSurprise's talk. As pointed out above, oh is often considered the most prototypical *surprise mark*, and it often is at least an element of these *surprise mark* LISTENER RESPONSES. Here, it is marked with a pitch rise-fall and the preceding click. It clearly is a LISTENER RESPONSE, given that Donna does not attempt to take over the floor and produces a *first assessment* as her next LISTENER RESPONSE.

Overall these examples show that similarly to *acknowledgements*, *surprise marks* are very brief and carry little lexical meaning. In contrast to *acknowledgements*, they are prosodically marked (albeit not strongly), often contain oh, are more sequentially restricted (post TCU, not mid-TCU), and tend to be followed by an elaboration on or rephrasing of the content marked as surprising. Relative to *first assessments*, *surprise marks* are less prosodically marked and briefer. The next LISTENER RESPONSE following it often is either an *assessment* or an *acknowledgement*. Just like *acknowledgements* they reinforce the Speaker-Listener-relationship and do not disrupt the flow of the ongoing multi-unit-turn.

Example 12 has the structure mentioned above: Donna follows up her surprise mark with a first assessment in line 5. We will now move on to discussing the sequential structure in which we find LISTENER RESPONSES that are first or second assessments, and attempt a characterisation.

125

5.4 Assessments

Assessments have been previously reported as potential LISTENER RESPONSES. In the present dataset, LISTENER RESPONSES that are *first assessments* tend to stand alone. Only sometimes does the main Speaker explicitly pick up on them to provide a second assessment before or while continuing their turn. LISTENER RESPONSES can also take the shape of second assessments. This phenomenon has not been described previously. In the following sections, I will present LISTENER RESPONSES that do *first assessment* and then LISTENER RESPONSES that do *second assessment* and characterise them with respect to their sequential position and their prosodic, lexical, and morphosyntactic form.

5.4.1 First Assessments

LISTENER RESPONSES that are *first assessments* are minimal, prosodically reduced relative to the surrounding talk or the current Listener's talk as a Speaker, and are not usually responded to by the main Speaker. They treat some aspect of the preceding talk as assessable and generally follow the prototypical structure of first assessments pointed out in the analysis of extract 5 (`[NP] + [BE] + [Adjective of Evaluation]`). Assessables can be a variety of things, ranging from a person, event, or thing being described to the gist of a longer stretch of talk. *First assessments* can be described in their sequential structure like this:

STRUCTURE OF FIRST ASSESSMENTS AS LISTENER RESPONSES

```
1    Speaker:    ongoing turn containing an assessable
2    Listener:   first assessment
3    Speaker:    continuation of ongoing turn
```

The first two examples are taken from the conversation between Angie and DaisyRae. Angie is talking about her son's insulin pump and the type of CGM he is using, called Enlite. She describes how they interface to help regulate his blood sugar levels. This set-up is called Smart Guard. DaisyRae provides frequent LISTENER RESPONSES, two of which are *assessments* – the first marked with an arrow in line 4, the second in line 25. We will look at them in turn.

(13) FIRST ASSESSMENTS, ANGIE AND DAISYRAE, SMARTGUARD, MINUTES 07:17-07:50

```
1        Angie:        and then it'll (.) it's got this thing called smart
2                      guard which can um (.)-
3        Angie:        [!] suspend the basal rate?=
4   →    DaisyRae:     =°°oh brillia[nt°°-
```

```
5       Angie:                        [so [if it knows that he's (.) going to
6                       (.) go low-
7       DaisyRae:                        [°°yeah°°.
8       Angie:          or if [he's trending low-
9       DaisyRae:             [°yeah°.
10      Angie:          then it'll suspend-
11      Angie:          (.4)
12      Angie:          the rate I think maybe (.) twe:nty minute[s or
13                      something before-
14      DaisyRae:                                               [°okay°-
15      Angie:          he actually hits hypo [so:-
16      DaisyRae:                             [°yeah°-
17      Angie:          and it it catch- (.)
18      Angie:          like we were really-
19      Angie:          (.4)
20      Angie:          not very sure about it,=
21      Angie:          =[well I think you know we'll still-
22      DaisyRae:        [°yeah°.
23      Angie:          give lucozade or sort of if we see the arrow's going
24                      down.
25      DaisyRae:       °°yeah°°=
26      Angie:          =but it does catch most of [them.
27 →    DaisyRae:                                  [°°that's great.°°=
28      Angie:          =yeah I think just if [he's had a lot of-
29      DaisyRae:                             [°°yeah°°.
30      Angie:          if he's got a lot of active insulin-
31      Angie:          so if he's got a lot of insulin on board,

                        ANGIE CONTINUES
```

DaisyRae treats Angie's description of the smart guard's functionality as an assessable. Angie completes the full turn and ends on a strong final rise (line 3). This can signal both continuation and an invitation to respond. Continuation is also projected by Angie's speech tempo: there is no indication of her slowing down, which would generally precede a handing-over of the floor. DaisyRae's *first assessment* oh brilliant (line 4) immediately latched on to the end of Angie's turn, and Angie continues in overlap with the final articulation of DaisyRae's Listener response to elaborate further on the Smart Guard.

The *first assessment* `oh brilliant` is prosodically reduced – it is produced with low amplitude, a narrow pitch range, and low pitch overall in comparison both to Angie's talk and to DaisyRae's behaviour as a Speaker. A further cue that this *first assessment* is a Listener response and not a claim to the floor is that DaisyRae produces *acknowledgements* in overlap with Angie's talk (`yeah` in lines 6 and 8). *Acknowledgements* are the most frequent, prototypical and unambiguous Listener Responses. They clearly frame this *first assessment* as 'one of their kind', not an attempt to take over the floor. This assessment is slightly different from the prototypical form described for example by Pomerantz (1984), but the form is recognisable in the underlying structure as `oh (that is) brilliant`, containing a noun phrase or deictic (that), a form of BE, and an adjective of evaluation (`[NP]` `+ [BE] + [Adjective of Evaluation]`).

Angie makes no explicit second assessment, treating DaisyRae's *first assessment* as unproblematic. She keeps the floor from lines 5-26, and DaisyRae continues to signal her Listenership with brief *acknowledgements*. DaisyRae produces the next *first assessment* in line 27, treating Angie's statement that the Smart Guard catches most hypos before they happen as an assessable. This first assessment overlaps with the last word of Angie's turn, a point at which completion is projectable. This assessment, like the first example we saw, is prosodically reduced – amplitude, pitch, and pitch range are lower than the surrounding talk – and it is lexically minimal. Its structure follows the familiar pattern `[NP] + [BE] +` `[Adjective of Evaluation]`. DaisyRae makes no grab for the floor; as soon as Angie continues speaking DaisyRae produces another *acknowledgement* in overlap, reaffirming her Listenership (line 29).

Angie's continuation is latched on to the assessing adjective `great`. She begins her turn with the acknowledgement token `yeah`. This pro-forma agreement leads up to a hedged downgraded disagreement or rather qualifying statement, noting that the Smart Guard really only works well in specific situations. The fact that Angie moves to do a second assessment to qualify DaisyRae's *first assessment* suggests an orientation to alignment as the default. If the main Speaker agrees with the Listener's first assessment, there is no need to produce a second (see also Donna's first assessment `that's good` on line 30 in extract 1). If the main Speaker does not align with the *first assessment* in some way, the assessment needs to be re-negotiated. Disagreements as in this case are marked as dispreferred responses, while upgrading agreements are treated as preferred responses.

To summarise, Listener Responses that are *first assessments* can but do not need to overlap with the ongoing turn, they can occur at points of possible completion or incompletion, and they treat an aspect of the preceding turn as assessable. They are generally optional, however they might follow a turn with rising pitch or

ending on a tag question. *First assessments* in this position are lower in amplitude, pitch, and pitch range, and minimal in their lexical realisation.

[Response cry] +	[NP] +	[BE] +	[EVALUATOR]	Extract
oh	∅	∅	brilliant	13
∅	that	's	good	12
∅	that	's	great	13

TABLE 5.1
SUMMARY CANONICAL STRUCTURE FIRST ASSESSMENTS

The *first assessments* analysed follow the sequential structure outlined in table 5.1. Response cries (Goffman 1978) are optional, and NP and BE can be dropped. A second assessment by the Speaker is only needed to modify the *first assessment* in some way in cases where there is disagreement or disalignment (on (dis)agreements see Pomerantz (1984), and on alignment vs affiliation Stivers (2008)); continuation of the ongoing project is the norm. The Listener often responds to the next increment produced by the Speaker with an *acknowledgement*, reaffirming their own status as a recipient.

5.4.2 Second Assessments

Listeners can not only produce *first assessments* in response to an assessable in the ongoing Speaker-talk, but also *second assessments* if the Speaker makes a first assessment of some aspect of their own talk. This has not been discussed previously in the literature, possibly because this category might have been subsumed under the general heading of 'assessments'.[4] *Second assessments* as LISTENER RESPONSES can follow the canonical format modelled on the preceding first assessment like `wow that is unusual` in extract 14, but they can be as minimal as `yeah` in extract 16. *Second assessments* tend to be prosodically more prominent than LISTENER RESPONSES that are *first assessments* or *acknowledgements*. The more minimal they are lexically, the more important pitch marking and interactional context become in distinguishing them from mere *acknowledgements* (see also Eiswirth (2014) on second assessments). Their brevity and their treatment as brief insertions that require no response other than Speaker-continuation show that they can be used as LISTENER RESPONSES. Their sequential structure can be characterised like this:

STRUCTURE OF SECOND ASSESSMENTS AS LISTENER RESPONSES

```
1   Speaker:    assessable
```

[4]In the inter coder reliability analysis we have seen that this distinction was challenging for a coder not trained in interactional analysis, and that second assessments are very rare overall. Hence, it is understandable why they might have been grouped together in the past.

```
2    Speaker:    first assessment (potentially implicit)
3    Listener:   second assessment (always preferred/aligning)
4    Speaker:    (receipt+) continuation
```

The examples presented will move from lexically complete to lexically minimal and prosodically marked. Extract 14 is taken from the conversation between Tess and Velominati. They have been discussing celebrities with Type 1 Diabetes, and Velominati introduces Theresa May as one of them (line 1). The Speaker-Listener assessment sequence is marked with arrows (lines 12 and 13).

(14) SECOND ASSESSMENTS, TESS AND VELOMINATI, T1D CELEBRITIES, MINUTES 27:50-28:00

```
1          Velominati:   and Theresa May our prime minister.
2          Tess:         ↑is she.
3          Velominati:   yeah.
4          Tess:         I didn't know that.
5          Velominati:   yeah: quite- [quite a recent diagnosis.
6          Tess:                      [< <allegro> °don't know if that's a link
7                        though°.>
8          Tess:         [that's very interesting.
9          Velominati:   [in- only last year     ] or something.
10         Tess:         and /type one,
11         Velominati:   at that age-
12 →       Velominati:   that's quite unu[sual isn't it.
13 →       Tess:                         [/wo:w that is unusual.
14         Velominati:   yeah.
15                       (1.1)
16         Tess:         wonder would it change her-
17         Velominati:   u::h-
18         Tess:         policies or mood or something,
19         Tess:         < <laughing> maybe that's been the issue here.>
```

TESS CONTINUES

Tess states that she did not know that Theresa May had T1D, and proffers a first assessment in line 8: that's very interesting. This first assessment thus seems to orient to the newness of the information *that* the current Prime Minister is a fellow person with Type 1 Diabetes, but not to the time of her diagnosis. Velominati,

on the other hand, orients to this as the surprising and assessable aspect.[5] Velominati expands on the late diagnosis and the question of age in lines 9-11. He then makes a first assessment orienting to these facts as assessables: `that's quite unusual` in line 12. As soon as `unusual` is projectable, Tess produces an upgraded agreement in overlap with Velominati's continuation., `wow that is unusual`. `wow` and `is` do the upgrading. Velominati signals receipt of the *second assessment* with a quiet `yeah`, then a pause of 1.1 second unfolds. The fact that Tess does not begin to speak for such a long time indicates that she does not treat her *second assessment* as closing-implicative; instead she behaves as though she expects Velominati to continue. This shows her orientation to him being the current main Speaker, and her *second assessment* being a LISTENER RESPONSE. When Velominati does not pick up the thread of the conversation for over 1 second, Tess takes the floor to continue with a topic-shift from diagnosis to policy. In this way, she orients to the assessment sequence as closing-implicative and making way for a change of topic. We will see an even more clear-cut example in the next example.

Extract 15 contains a *second assessment* which follows a similar structure. It is taken from the conversation between Darren and Angie. Darren has just talked about the trouble he had in school before and after his diabetes diagnosis. Angie has asked him if things are better now, and Darren says that he now has friends at work, at university, and through diabetes-related activities. This is what Angie's first assessment at the start of this extract refers to:

(15) SECOND ASSESSMENTS, ANGIE AND DARREN, FRIENDS WITH DIABETES, MINUTES 08:15-08:25

```
1       Angie:      it's good to- it's good to have some friends that-
2       Angie:      kno::w and understa:nd exactly what'-s [involved and
3                   just=
4  →    Darren:                                        [yeah it's nice.
5       Angie:      =to sound off about things and they-
6       Angie:      they completely get where [you're coming from [yeah.
7       Darren:                               [°yeah.            [it's
8                   nice.
9       Darren:     cause usually if you spoke to your normal friends
10                  they're like-
11      Darren:     'what',
```

ALTERNATING SHORT TURNS

[5]T1D is also referred to as 'Juvenile Diabetes' and the majority of new diagnoses are children. Thus, such a late diagnosis is indeed unusual.

Angie's first assessment follows the canonical structure outlined above (`it's good [...]`), and then she expands on what exactly is good about having friends with the same condition, and why. When a first point of possible completion is projectable, Darren utters an *agreement* in overlap with Angie's ongoing turn. Angie's assessment is at a point of possible syntactic completion after `understa:nd exactly`, and also after `involved` (line 2), and `involved` is precisely where Darren's agreement comes in. Instead of stopping Angie rushes through to further elaborate her assessment in lines 5 and 6, thereby rendering Darren's insert as a Listener response.

The point at which possible syntactic completion is projectable comes in line 6, when Angie states `they completely get where you're coming from`. This formulaic expression is projectable at `where`, and Darren proffers a `yeah` in overlap. At the end of the increment, Angie appends a `yeah`, and Darren repeats his previous *second assessment* in overlap with this final tag-yeah. He then takes the floor, continuing the project Angie has started – elaborating on why it is good to have friends with the same condition. These two instances of a nearly identical realisation of a *second assessment* show how fine the line between Listener response or not can be: in the first case, the main Speaker projects continuation and does indeed continue her turn, treating Darren as a Listener, not as the next Speaker. In the second case, she completes the increment and Darren does take over the floor, making the second uttering of this second assessment in lines 7 and 8 an instance of incipient Speakership and not a Listener response. This example nicely illustrates that assessment sequences can be and in fact often are closing-implicative. They are used to prepare a topic-shift by closing off one topic (here, diabetic friends) and opening the floor for something new (Antaki et al. 2000; Goodwin and Goodwin 1987b; Lindström and Heinemann 2009; Mondada 2009).

These examples have shown how important sequential and interactional context are in determining whether an utterance is a Listener response or not. Similarly, the line between *acknowledgements* and *second assessments* can be blurry when the lexical format is not canonical. The next extract which contains several examples is taken from the conversation between Tomek and Velominati. The two men have been discussing eating out or attending social events as potentially challenging when following a very low carbohydrate diet, with Christmas Dinners as a specific example. Velominati has described the conversations he has had with friends, and the beginning of this extract concludes this.

(16) SECOND ASSESSMENTS, TOMEK AND VELOMINATI, SUPPORTIVE FRIENDS, MINUTES 06:48-07:15

```
1       Velominati:    an:d they became quite interested in it.
```

132

```
2      Velominati:     [you know (.) rather than being something that-

3      Tomek:          [°mhm°.

4                      (.6)

5      Velominati:     wouldn't be spoken about in [polite conversation.

6      Tomek:                                      [°yeah°.

7      Velominati:     it became an issue that [people were-

8      Tomek:                                  [°mm hm°.

9      Velominati:     quite happy to- [to talk about and discuss.

10     Tomek:                          [°to discuss°°°yeah°°.

11     Tomek:          well ↑I think that's very ↑good,

12     Tomek:          ↑isn't it,

13     Tomek:          [you know that people are sort of very open about it

14                     and, you know, understand your needs and want to-

15  →  Velominati:     [↑oh yeah:-

16     Tomek:          help you (.) you know (.) to um (.) you know-

17     Tomek:          to manage your condition [better.

18  →  Velominati:                              [/yes\.

19  →  Velominati:     yeah yeah (.) [/that's right\.

20     Tomek:                        [by even being understanding of,

21     Tomek:          you know of of of that con[dition and and and-

22     Velominati:                               [uh-huh-

23     Tomek:          the needs that you require.
```

TOMEK CONTINUES

This extract contains three *agreements* with slightly different first assessments presented by Tomek. From lines 1-9, Velominati describes his friends' attitude to his diet and Tomek provides minimal *acknowledgements* in lines 3, 6, and 8. When Velominati comes to a point of possible completion, Tomek takes the floor with a summary assessment on line 11. He adds a tag `isn't it` (line 12). If we consider assessment sequences as adjacency pairs, a fitted response to this turn would be a second assessment. And indeed, Velominati agrees by saying `oh yeah:-` (line 15). There is a marked pitch rise-fall on `oh yeah`, marking it as different from a mere *acknowledgement*: these tend to be prosodically subordinate to the main-Speaker talk, with little to no pitch movement and low amplitude.

The response looks removed from this first assessment in the transcript because it occurs in overlap with Tomek's continuation and expansion of his assessment to a slightly different aspect of what is good about people being happy to talk about the diet. In lines 13-17 Tomek changes the scope of the positive

assessment initiated on line 11 to it being good that people understand and want to help. Just as above, a fitted response to a first assessment would be a *second assessment* of some sort. Velominati begins a tripartite agreeing response represented on lines 18 and 19: `yes.  yeah yeah (.)  that's right.` `yeah yeah` is uttered while Tomek takes a breath, and the third part `that's right` overlaps with the beginning of Tomek's continuation. Both `yes` and `that's right` are marked with rise-falling pitch contours that mark them as more than *acknowledgements*. Following this, Velominati makes no attempt at taking the floor, thereby confirming his Listener status, and Tomek continues for another 9 seconds.

I hope to have shown that *second assessments* as Listener Responses take different shapes depending on their design: the less they are elaborated lexically, the more they will be marked prosodically to distinguish them from mere *acknowledgements*. *Second assessment* Listener Responses are different from the usual sequential pattern of assessment sequences in that the main Speaker produces both the assessable and the first assessment. They might also utter an assessment, for example at the end of a narrative, that orients to the resolution or a specific aspect of the preceding talk. Such an assessment seems to make relevant more than mere acknowledgement (similar to what we have seen in example 9). *Second assessments* that are Listener Responses must be the preferred response because a dispreferred second would require more elaboration than possible in this interactional slot (see Pomerantz (1984) on preference organisation).

5.5   Repair

There is very little mention of *repair* overall in the existing literature on Listener Responses, despite Schegloff (1982) arguing that *repair-initiation* is one of the few next activities that are always potentially relevant in a conversation. He then focuses his argument on *acknowledgements* as explicitly passing on the opportunity to initiate repair. However, the point remains true: Listeners could *initiate repair* just as well as do an *acknowledgement*. Gardner (2001) does make reference to repair-initiation in his work, but there is little beyond this mention. I have not found a discussion of Listeners actually doing repair. Both, however, happen in this collection.

Listeners very rarely *do repair* because Speakers rarely signal trouble finding a word or formulation. However, if the main Speaker indicates trouble, the Listener can provide a candidate repair item. I will present some examples for this type of repair first. Listeners can also *initiate repair* by asking the main Speaker more or less directly to elaborate on a specific aspect of the preceding talk. This

other-initiation of repair is more frequent than Listeners doing repair and will be discussed in the second subsection.

5.5.1 Self-initiated Other-repair (word-search completions)

There are a few instances in which the Speaker indicates trouble finding a word or phrases needed to continue their turn. Cues can be filled or unfilled pauses and hesitations, or repeated restarts and hesitation markers. The Listener can then come in to provide a candidate continuation. This repair is generally lower in amplitude than the surrounding talk and not marked with a pitch accent, thereby prosodically marking it as not an attempt to take over the floor. In the data analysed here, the only pitch movement observed on an item doing repair was a slight rise, cueing it as tentative, merely a suggestion. In many cases repair and continuation overlap and the repair is not separately ratified by the main Speaker, in other cases it is included in their continuation as though it was their own. This pattern can be generalised as follows:

STRUCTURE OF SELF-INITIATED OTHER-REPAIR AS A LISTENER RESPONSE

```
1    Speaker    : initiates repair in ongoing turn
2    Listener   : repair (word supply)
3    Speaker    : (ratification +) continuation
4    Listener   : acknowledgement in overlap with continuation
```

The first example of *self-initiated other repair* is taken from Tomek and Velominati's interaction (extract 17). They are discussing very-low carb versus low glycaemic index diets and Tomek claims he does not need insulin for certain foods that have a low glycaemic index.

(17) SELF-INITIATED OTHER-REPAIR, TOMEK AND VELOMINIATI, QUINOA, MINUTES 16:10-16:20

```
1         Tomek:        so:: uh things like (.) uh-
2         Tomek:        (.4)
3         Tomek:        quini- qui- uh no not quini- qu- ↑quoina (.) quina,
4    →    Velominati:   °er quinoa°,
5         Tomek:        yeah that's the [one sorry-
6         Velominati:                   [°yeah°-
7         Tomek:        (.7)
8         Tomek:        and uh-

                        TOMEK CONTINUES
```

Tomek signals trouble finding and pronouncing the target word. He first pauses, then produces the hesitation marker uh, pauses again, and then produces several false starts. He then pronounces quinoa in a number of different ways, with increasingly marked rising intonation. These are clear signals of repair-initiation, and Velominati eventually does the *self-initiated other-repair* by producing the target pronunciation of quinoa with lower amplitude and flat pitch. Tomek ratifies this repair in the next turn, excusing his mispronunciations. Velominati produces an *acknowledgement* in overlap with Tomek's ratification. He makes no move to take the floor during the .9 second pause before Tomek continues with the next TCU. Both show a clear orientation to the relevance of Tomek continuing – Velominati by remaining a Listener and not producing further talk, and Tomek by continuing after a brief hesitation.

The second example, extract 18 also relates to ketogenic diets. The instance of *self-initiated other-repair* is at the very end of this extract (line 26), but we will return to the first lines in the next section on other-initiation. Connor has just told PuzzleB that he regards food as medicine, with the ketogenic diet having made his diabetes management much easier.

(18) SELF-INITIATED OTHER-REPAIR, CONNOR AND PUZZLEB, KETOGENIC DIET, MIN-
      UTES 07:10-07:40

```
1       Connor:      I feel I kinda manage my diabetes quite (.) well now,
2       Connor:      u[m::::-
3       PuzzleB:       [mhm-
4       Connor:      you know (.) that's it (.) I don't know-
5       Connor:      I think-
6       Connor:      ((inbreath 1.2))
7       Connor:      you know I think I kinda like um-
8       Connor:      like I'm on like a like-
9       Connor:      a kinda ketogenic diet and things like that-
10      Connor:      a lot of the things I learned-
11      Connor:      [is through-
12      PuzzleB:     [a what diet-
13      Connor:      a ketogenic diet,
14      PuzzleB:     no Ive not [heard of that.
15      Connor:                 [oh no oh      [< <allegretto> it's like a
16                  very very lo:w ca:rb>-
17      PuzzleB:                              [right-
18      PuzzleB:     right-
19      Connor:      like um it kinda's like-
```

```
20        Connor:      kinda hi- high fat and protein diet,
21        PuzzleB:     [right.
22        Connor:      [and I feel like it keeps my blood sugars-
23        Connor:      (.7)
24        Connor:      like-
25        Connor:      (.4)
26        Connor:      s- ve- [just stable and I feel [great-
27  →     PuzzleB:            [°steady°-
28        PuzzleB:                                     [right-
29        Connor:      I mean I can go out for an exercise and I go out like-
```

CONNOR CONTINUES

Connor has given a broad definition of ketogenic diets and moves on to describing its effects in line 22. The TCU he launches remains incomplete, continuation with an adjective is relevant. Connor first hesitates for .7 seconds, then produces the hesitation marker `like`, pauses again for .4 seconds, and then launches his continuation with two false starts. These false starts are beginnings of possible words, `s- ve-`, which restrict the set of potential continuations. In response to these possible repair-initiations PuzzleB then picks up Connor's false starts and offers `steady` as a possible continuation in line 27 in overlap with Connor's continuation. The repair's brevity and its prosodic design clearly mark it as a Listener response. It is quiet and carries no pitch accent. Connor continues in overlap, producing a slightly different lexical item `stable`. The two terms describe the same effect, and Connor does not further respond to PuzzleB's repair. PuzzleB acknowledges Connor's choice of term in overlap with the end of his TCU, and Connor then continues to elaborate how this stability impacts on his life. Here, just like in the first example, the word-search-trouble was clearly signalled by the main Speaker, and the *self-initiated other-repair* was minimal and unobtrusive.

5.5.2 Other-initiation (requests for information)

Listeners can also initiate repair – this is referred to as *other-initiation*. Other-initiation is based on some trouble with the preceding talk; this can be an issue with hearing or understanding, an orientation to the preceding as incomplete, or even (only once in the whole collection) wrong. Thus when speaking of *other-initiation* here we are not talking of corrections but rather *requests for information*. These repair-initiations lead to insertion sequences of different lengths (most fairly short) in which the Speaker does the required repair and then moves back to the project that has been momentarily suspended. However, I argue we can still consider

137

these actions Listener Responses because they make further same-Speaker-talk relevant and reinforce the current Speaker-Listener relationship. Effectively, they support the continuation of the ongoing project by making sure the Listener has all the details they need in order for their mind to 'stay with' the Speaker. The structure of *other-initiations* can be summarised as follows:

Structure of Other-Initiation as a Listener Response

```
1   Speaker:    repairable
2   Listener:   repair-initiation - request for information
3   Speaker:    provide information and continue
```

We have already seen two examples of other-initiation in extract 5 from the conversation between Lavina and Samantha. I would like to briefly refer back to them. The first one is Samantha's clarification question you're a pump user on line 8.

(19) Other-initiation, lines 5-15 from extract 5, Lavina and Samantha, Technology and DAFNE, minutes 02:30-04:00

```
5       Lavina:     so I ha::ve the dexcom.
6                   (.5)
7       Samantha:   °right°.
8       Lavina:     [!] um:-
9       Lavina:     (.7)
10      Lavina:     I:: started [using-
11 →    Samantha:              [you're a pump user.
12      Lavina:     yes.
13      Samantha:   yes.
14      Samantha:   °°mhm°°.
15      Lavina:     but I wasn't (.) whe:n I got the libre system.
```

LAVINA CONTINUES

Here, I would only like to highlight that the insertion sequence triggered by the *other-initiation* is minimal: Lavina confirms that she is indeed a pump user with a yes on line 9, Samantha receipts this by repeating yes, and then provides an acknowledgement mhm, upoon which Lavina continues with her ongoing project.

Extract 20 shows a rather extensive insertion following a *repair-initiation* by the Listener. It is the beginning of extract 18 presented in the previous section.

(20) Other-initiation, lines 8-16 from extract 18, Connor and PuzzleB, Ketogenic Diet, minutes 07:10-07:40

138

```
8       Connor:     like I'm on like a like-
9       Connor:     a kinda ketogenic diet and things like that-
10      Connor:     a lot of the things I learned-
11      Connor:     [is through-
12 →    PuzzleB:    [a what diet-
13      Connor:     a ketogenic diet,
14      PuzzleB:    no I've not [heard of that.
15      Connor:                 [oh no oh     [< <allegretto> it's like a
16                  very very lo:w ca:rb>-
17      PuzzleB:                             [right-
```

CONNOR CONTINUES

Connor introduces the topic of his ketogenic diet (line 9), but before he can elaborate on the things he has learned (lines 10/11) PuzzleB initiates repair by asking for clarification: a what diet (line 12). This *repair-initiation* could refer either to a problem in hearing or to a lack of knowledge about what a ketogenic diet is. Connor interrupts his ongoing turn when PuzzleB initiates repair and repeats the term. This addresses any potential problems in hearing. PuzzleB responds by making her lack of knowledge explicit: no I've not heard of that (line 14). Connor begins addressing the trouble-source, the lack of information, in overlap with this turn. PuzzleB reaffirms her Listenership with an *acknowledgement* in the last line given here. Once Connor has broadly outlined the idea of a ketogenic diet he moves back to his main project, arguing that he does not need his diabetes team very much because this way of eating allows him to keep his glucose levels in a fairly narrow range, and to exercise regularly but also spontaneously. Thus, even though the *repair-initiation* triggers a brief insertion, it still orients to the Speaker-Listener role distribution and ensures that the bigger project the participants have embarked on can continue.

The final example I would like to refer back to comes from the end of extract 5 and demonstrates that *other-initiation* makes a different next action relevant than an *acknowledgement*.

(21)   OTHER-INITIATION, LINES 83-90 FROM EXTRACT 5, LAVINA AND SAMANTHA, TECH-
        NOLOGY AND DAFNE, MINUTES 02:30-04:00

```
83      Lavina:     and at that stage I also had the joy of a marked dawn
84                  phe↑nomenon,
85 →    Samantha:   °a wha-°/yes-
86      Lavina:     so I:: would be up in-
```

139

```
87      Lavina:        twenty-two twenty-[three in the morning,
88      Samantha:                        [but that's also the compensation=
89      Samantha:      =for the hypo in the [middle of the night.
90      Lavina:                             [yes.
91      Lavina:        ex[actly.
92      Samantha:        [°mhm°.
```

LAVINA CONTINUES

As mentioned in the initial discussion of the extract as a whole, at the end of this extract Lavina brings up her Dawn Phenomenon (line 82,83). This is a technical term, and Samantha begins to *initiate repair* with a `wha-` in line 84. This might be signalling a problem in hearing or in understanding, either with respect to the meaning or the relevance of Lavina's preceding statement. However, she self-corrects, effectively repairing her own repair initiation, by cutting it off mid-way. She replaces it with an *acknowledgement*, `yes`. The cut-off *repair-initiation* is quieter than the surrounding talk, and the following *acknowledgement* has a higher amplitude than the other Listener Responses Samantha has produced so far in the extract. This contrastive stress emphasises the *acknowledgement* and cues that it cancels out the *repair-initiation* – and indeed, Lavina does not do repair in the following turns. Instead, Samantha in lines 87 and 88 shows that she had interpreted the high morning readings as the body compensating for the night time hypoglycaemias. This also explains what Samantha had been orienting to as a repairable – not trouble in hearing or in knowledge, but rather in relating the morning highs to the Dawn Phenomenon instead of seeing them as a physiological response to the hypoglycaemias at night. This restart demonstrates that *repair-initiations* and *acknowledgements* make different next actions relevant; the ensuing repair-sequence momentarily suspends the ongoing project but still maintains the Listener-Speaker role distribution, while an *acknowledgement* signals no trouble and invites immediate continuation.

To summarise, we can say that these *other-initiations* tend to be similar in amplitude to the surrounding talk and often carry a pitch accent. There usually is overlap between the ongoing Speaker-turn and the *repair-initiation*, as well as between initiation and the actual repair. The insertion sequence triggered by the *repair-initiation* can be very brief. The Speaker can even continue without addressing it explicitly if what has been marked as repairable is resolved in the immediately following talk. It can also be more extensive, as in example 20. Speakers tend to mark the transition back to their momentarily suspended ongoing project for example through format-tying or with brief pauses. Even though the

repair-initiations trigger an insertion sequence, the Speaker-Listener-relationship is never put into question; the Listener only takes a more directive role than with the other LISTENER RESPONSES we have seen so far.

5.6   Collaborative Completion and Voicing

On a level of involvement quite similar to *other-initiation*, but without steering the conversation as actively, Listeners can also *co-construct* stretches of talk together with the main Speaker. They can either do this within stories, by *voicing* a character in the main Speaker's story, or by *collaboratively completing* the other's ongoing turn. As Wilkinson and Kitzinger (2014: 148) nicely demonstrate, the turn that is being completed by the Listener is oriented to by both parties as the 'Speaker's turn'. I will briefly refer back to examples for both phenomena that we have seen in earlier extracts. We will start with two examples where the Listener *voices* one of the characters in a narrator's story. Listeners usually follow up their unusually long contribution with an *acknowledgement* or other minimal response, signalling the Listener-Speaker-roles are not in question. This is in line with Wilkinson and Kitzinger's (2014) observations. Structurally, we can represent *voicing* as a Listener response like this:

STRUCTURE OF VOICING AS A LISTENER RESPONSE

```
1    Speaker:    character in story speaks, response is projectable
2    Listener:   [voicing
3    Speaker:    [continuation OR explicit reaction to voicing
4    Listener:   acknowledgement
5    Speaker:    continuation
```

The two examples I will draw on are from the first two long extracts: Donna's *voicing* of what PuzzleB might have said to her daughter (line 10) when she was on the train to the diabetes camp, with her blood sugar levels soaring because of the excitement, and Samantha's *voicing* of what Lavina might have said to her diabetes nurses:

(22)   VOICING, LINES 7-14 FROM EXTRACT 1, DONNA AND PUZZLEB, DIABETES CAMP AND DAFNE, MINUTES 16:15-18:00

```
 7        PuzzleB:        'My °blood's°< <laughing> twenty- [↑fou:r,'>
 8        Donna:                                           [ha
 9        PuzzleB:        I'm < <laughing> 'Oh my /go:d,'>
10  →     Donna:          [< <laughing> °come /ba:ck°->
```

```
11       PuzzleB:       [< <laughing>I think the stress was kicking in=

12       PuzzleB:       =he he he[he he->

13       Donna:                  [he he he-

14       PuzzleB:       and then there was no reception after that for a while-
```

PUZZLEB CONTINUES

Donna's *voicing* of PuzzleB's likely response in the story overlaps with PuzzleB's actual continuation that moves the story forward. The *voicing* is produced in matched pitch and rhythm, and followed up with laughter in overlap with PuzzleB's laughter, and PuzzleB continues her narrative.

The second example I would like to refer back to was presented in extract 5. Lavina told Samantha about her DAFNE experience and how her diabetes team suggested she get up several times every night to check her blood sugar. Samantha then *voices* a likely answer.

(23)   VOICING, LINES 46-52 FROM EXTRACT 5, LAVINA AND SAMANTHA, TECHNOLOGY
        AND DAFNE, MINUTES 02:30-04:00

```
46       Lavina:       'can you get yourself up at two o'clock in the

47                     morning?'

48  →    Samantha:     'no I can't than[k you [I've got a job to go to.'

49       Lavina:                       [!]    [no I'd-

50       Samantha:     [((laughing))

51       Lavina:       [I- did do tha[t-

52       Samantha:                   [°°mhm°°-
```

LAVINA CONTINUES

Just as in the previous example, the *voicing* prosodically matches the preceding turn, blending into the narrative and marking the contribution as part of the ongoing project, not something that would interrupt or distract from it. Here, however, Lavina does comment on it explicitly because Samantha's suggested *voicing* is not what Lavina actually said. Thus, Lavina initiates repair. Samantha remains a Listener, signalling her continued recipiency with a laugh in line 50 and `mhm` in line 52.

As stated in the introduction of this section, Listeners can also complete the Speaker's sentence or utter some words alongside the main Speaker's talk. These *collaborative completions* tend to connect to TRPs where continuation is highly projectable. Usually the Speaker in fact continues in overlap with the Listener. The Listener then produces an *acknowledgement* as soon as their contribution is

complete. They thereby signal continued Listenership in an unambiguous manner at the next possible point, and the Speaker continues their project. Structurally, this can be summarised like this:

STRUCTURE OF COLLABORATIVE COMPLETION AS A LISTENER RESPONSE

```
1    Speaker:    highly projectable turn
2    Listener:   [collaborative completion
3    Speaker:     [continuation OR reaction to completion
4    Listener:   acknowledgement
5    Speaker:    continuation
```

The first example I'd like to refer back to is an earlier point in the conversation between Lavina and Samantha, where Lavina introduces the topic of the DAFNE course.

(24)   COLLABORATIVE COMPLETION, LINES 21-42 FROM EXTRACT 5, LAVINA AND SAMAN-
       THA, TECHNOLOGY AND DAFNE, MINUTES 02:30-04:00

```
21         Lavina:       I:: did a DAFNE-
22         Lavina:       the DAFNE [training course in ↑f::ebruary twenty ten-
23         Samantha:              [/yes-
24         Samantha:     °°mhm°°.
25  →      Lavina:       a::nd (.) that highlighted what I'd been-
26         Lavina:       telling my diabetes team for twenty odd years.
                         SIX LINES OMITTED

33         Lavina:       but it wasn't until I did DAFNE-
34         Lavina:       (.7)
35         Lavina:       that straight away-
36  ⇒      Samantha:     [it highlighted it.
37         Lavina:       [when my readings were put [on the board-
38         Samantha:                                [°°mhm°°.
39         Lavina:       the nurses that I'd been seeing for twenty-odd years
40                       said-
41         Lavina:       'yes, we think you're having-
42         Lavina:       ↑nighttime [hypos.'

                         LAVINA CONTINUES
```

In line 25, Lavina mentions the course `highlighted` an issue she had been telling her diabetes team about for years (see single arrow). She returns to this

point in line 33. Samantha recycles her formulation, *completing* Lavina's turn with `it highlighted it.` (line 36, double arrow). Lavina, however, continues in overlap with Samantha's *candidate completion*, describing the scene further. Samantha reinstates her Listenership with an overlapping `mhm` as soon as she has finished her contribution. The second example, taken from extract 11, is Tess *completing* one of Velominati's turns.

(25) COLLABORATIVE COMPLETION, CONTINUATION OF EXTRACT 11, TESS AND VELOMI-
      NATI, URINE ANALYSIS, MINUTES 2:55-3:10

```
10         Velominati:      [!] so the only way of self-monitoring was uh-
11         Velominati:      the delightful-
12         Velominati:      urine an[alysis.
13  →      Tess:                    [urine dip.
14         Velominati:      [yeah (.) you know-
15         Tess:            [yeah.
16         Velominati:                            [with a little test tube and some-
17         Tess:                                  [yeah-
```

Velominati describes how when he was diagnosed there was no simple blood glucose monitoring – he still had to use urine samples to get a rough indication of his glucose levels. As he moves on to utter the term `urine analysis` (line 12), Tess *completes* his turn with `urine dip` (line 13). Velominati receipts this with a `yeah` and immediately continues to describe the details of how the urine analysis was done. Tess responds, reaffirming her Listenership, with repeated *acknowledgements* in overlap with Velominati's talk (lines 15 and 17).

These examples demonstrate that *collaborative completions* and *voicing* within narratives can also be LISTENER RESPONSES. Even though they are longer, more involved, and rarely prosodically reduced, the interactants negotiate their status as LISTENER RESPONSES on the spot. The Speaker usually continues in overlap with the Listener's contribution, and the Listener produces an *acknowledgement* almost as soon as they have finished their contribution, thereby reaffirming their role as a Listener. Because they are both fairly rare and so similar in structure, I treat them as a shared category of *joint utterances* in the next chapters.

5.7 Discussion

Listeners can respond to ongoing other-Speaker-talk with a number of actions that do not disrupt the flow of the ongoing multi-unit-turn, lay no claim to the floor, and do not put the current Speaker-Listener-relationship into question. All LISTENER

Responses further share an orientation to the preceding talk, in the sense that different kinds of Speaker-TCUs make different Listener Responses relevant. The summary in table 5.2 shows how Speaker-TCUs preceding and following a Listener response systematically differ in what they contain or do. At the same time, the Listener response also orients to the preceding Speaker-TCU in a specific way; in the case of a 'neutral' TCU treating it either as unproblematic (*acknowledgement*) or as repairable (*other-initiation*). *Acknowledgements* are the most well-researched and also the most frequent type of Listener Responses. They can be considered the prototypical Listener response, but they are certainly not the only one.

This analysis and the summary in table 5.2 also suggest there are three types of Listener Responses with respect to how much they are constrained by the preceding talk, and thus the Speaker's actions:

1. other *repair-initiation* (requests for information) and *acknowledgements* are always potentially relevant in the sense that the Listener can very freely decide when to produce them;

2. *self-initiated other-repair* (word supplies) and *second assessments* are, as their name suggests, things the Speaker needs to prepare by producing the relevant first part, and the Listener then responds to accordingly;

3. the other action types – *surprise marks, joint utterances (collaborative completion and voicing)*, and *first assessments* are negotiated between Speaker and Listener and not clearly dominated by one party in terms of where and when they (can) occur.

This distinction is interesting and warrants closer qualitative analysis. It will also be relevant and I will elaborate on it further in the quantitative analysis in chapter 6, where I look into gender-related variation in the distribution of the different *actions* Listeners do.

This paper contributes to our holistic understanding of Listener Responses by discussing *second assessments* and *other-initiation* as actions Listeners can do and remain Listeners. These have been analysed in various other contexts so far, but not as Listener Responses. I have attempted to show here how longer and more involved Listener Responses are marked as not taking over the floor. It would be a worthy endeavour to focus on Speakership incipiency of all possible Listener response *actions*, similar to the work that has been done on *acknowledgements*, and to summarise which tools and strategies interactants use to cue whether they are doing being a Listener, or making a move to take the floor.

Acknowledgements, surprise marks, and minimal *second assessments* or *other-initiations* draw from the same lexical material, but rarely does a participant mistake

one for the other. Interactionalists use prosody in addition to the exact sequential placement as contextualisation cues to negotiate which action for example a `yeah` is doing: a drawn-out, pitch-marked variant that responds to a first assessment can be understood to do *second assessment*, while a short `yeah` with low amplitude and no pitch mark in any other place in the other's multi-unit-turn tends to be treated as a mere *acknowledgement*. There is also some overlap between *surprise marks* and *first assessments*, given that participants often treat surprising or new information as assessable. *Surprise marks* are most often combined with `oh`, while *first assessments* either follow or can be derived from the prototypical assessment structure of `[NP]+[BE]+[Evaluator]`. I will delve further into this complex form-function relationship in chapter 7.

I hope to have shown three things in this analysis: first of all, that there is a cohesive set of LISTENER RESPONSE *actions* that goes beyond mere *acknowledgement*. Second, that all of them are things Listeners can do without taking over the floor or disrupting the Speaker's ongoing project, and third that they can be distinguished based on the exact sequential context. I have further pointed out the importance of prosody in distinguishing between different LISTENER RESPONSE *actions*. As indicated above, we will return to this in the last analysis chapter.

Such a taxonomy is interesting from a CA point of view to survey all the possible actions Listeners can do while remaining Listeners, but it has even more potential as the basis for a coding scheme that allows for quantitative analyses of variation in LISTENER RESPONSES. These analyses can be done both with respect to their frequency and with respect to which actions get done when in an interaction, and in which ways (using which lexical material, or specific prosodic design). This is the focus of the next two analysis chapters – an analysis of variation in the LISTENER RESPONSE *actions* male and female Listeners and Speakers do, followed by an exploration of a potential form-function relationship for the different Action Types.

	Acknowledgement	Surprise Mark	First Assessment	Second Assessment	Self-Initiated Other-Repair	Other-Initiation	Collaborative Completion or Voicing
1. Speaker			**Ongoing Multi-Unit-Turn**				
Characteristics of TCU	Ongoing Multi-Unit-Turn	contains something surprising	contains an assessable	contains first assessment	contains repair-initiation	contains re-pairable	continuation highly predictable
2. Listener			**Listener Response**				
Lexical Characteristics	minimal	minimal	[NP] + [BE] + [Evaluator]	minimal	single word	question-format	syntactically matched
Prosodic Characteristics	prosodically reduced	can be pitch marked	can be pitch marked	usually prosodically reduced	prosodically reduced	prosodically matched	prosodically matched
2a. Speaker **2b. Listener**						Repair Receipt or Ratification	
3a. Speaker (optional)			2nd Assessment	Receipt	Receipt or Ratification	Receipt	Receipt
3. Speaker			**Continuation of Multi-Unit-Turn**				
Characteristics of continuation TCU	new information or elaboration	elaboration on surprising item	potentially second assessment	new information	elaboration	elaboration	elaboration

TABLE 5.2

OVERVIEW LISTENER RESPONSES ACTION TYPES

Chapter 6

Who does what, and how often? Operationalising Action Type Frequency

> [Listener Responses] are fitted to the details of the locally preceding talk, and cannot be properly understood or appreciated when disengaging from it.

Emmanuel Schegloff (1982: 86)

This chapter presents the third step in integrating interactional and variationist methods and shows that there are important interactional constraints on variation in ACTION TYPES: Similar to overall frequency, *which* actions get done depends on both the Speaker's and the Listener's behaviour.

It contributes to sociolinguistic theory and methodology in three crucial ways: By proposing an interactionally sensitive way of (1) quantifying, and (2) interpreting variation in LISTENER RESPONSE ACTIONS, and (3) introducing Bayesian Zero-Inflated Poisson (ZIP) regression models as a tool that allows us to go beyond the distributional approach and take random effects into account when analysing within-category variation. The classification is based on the taxonomy of Action Types developed in chapter 5, and the applicability of this coding scheme is underlined by the exceptionally high inter coder reliability (see chapter 3).

In the present chapter, I conceptualise LISTENER RESPONSE ACTIONS as the variable and the *individual Action Types* as the variants, broken down at different levels of granularity for the three models.[1] Variation in their frequency distribution is analysed based on the social variable gender.

---

[1]As in the previous chapters, the VARIABLE is referred to in SMALL CAPS, and the *variants* are *italicised*.

Descriptive statistics show that the female Listeners in my data produce more diverse Responses than male Listeners. With respect to Speaker gender, Listeners of either gender produce more diverse Responses towards male than female Speakers in the conversations analysed here. Specifically, women produce more *joint constructions* and *first assessments* than men, while male Listeners *initiate repair (ask questions)* relatively more frequently in the present interactions (for readability I will primarily refer to them as 'questions' in this chapter). The regression models reflect these effects: men produce fewer *first assessments* and *joint constructions* but more *questions* than women. Men as Speakers receive more *first assessments* and *questions*, and fewer *second assessments*. In the final model where ACTION TYPE is aggregated by *Main Actor*, actions driven by the Speaker are indeed more strongly influenced by Speaker gender, and actions not constrained by the Speaker (and thus driven by the Listener) by Listener gender. In all models, the effect size of Listener gender alone is a stronger and more consistent than that of Speaker gender.

It is important to note again that the contribution of my thesis does not focus on making any claims about gender-based variation in LISTENER RESPONSES beyond the present sample. Rather, I use this social variable to illustrate my methodological and theoretical points. What I demonstrate here is that 'meaning' and 'actions' are jointly produced, hence it is not only reductive but potentially misrepresents what is happening in the interaction to ascribe any action to one person (or group) only. An interactional perspective allows us to model the interactional constraints and understand what the variation observed actually means.

This chapter is structured as follows: I first present a distributional analysis of variation in LISTENER RESPONSE ACTIONS, followed by an interim discussion of how using the next-turn proof-procedure changes our understanding of the variation we observe. I then present Bayesian ZIP models that allow for the inclusion of random effects.[2] The discussion focuses on the theoretical and methodological implications of the analysis.

6.1 Two approaches to analysing variation in Action Types

In the following two sections, I first present descriptive statistics of the variation in ACTION TYPES between male and female Listeners and Speakers. Given the limitations of descriptive statistics, I then present ZIP models that take into account variation caused by the individual Listeners and Dyads. These models show that

---

[2]Within the frequentist framework, a model of this complexity would not converge. Furthermore, the Bayesian paradigm allows for the inclusion of random slopes *and* intercepts, which is conceptually important here.

the trends observed in the distributional analysis persist even when including random slopes and intercepts in a ZIP model.

6.1.1 Relative Frequency

In the analysis of how *often* Listeners respond in chapter 4, we have seen two things in this data: Firstly, in the recordings analysed, women overall respond more frequently than men. Secondly, we have seen cross-gender accommodation, with men slightly increasing their rate when Listening to female Speakers, and women slightly decreasing theirs when Listening to men. One of the central criticism of such accounts of overall frequency put forward by Conversation Analysts is that not all LISTENER RESPONSES are doing the same action (Schegloff 1982; Goodwin 1984; Drummond and Hopper 1993a). Consequently, even if men and women responded at the same or similar frequencies, they might still be doing different things. This frames LISTENER RESPONSES as a variable, and the *individual Action Types* as its variants. A taxonomy of ACTION TYPES based on a close analysis of the over 5000 Listener Responses in the present set of interactions has been presented in the previous chapter and forms the basis of the following analysis.

A common way of analysing variation in the frequency of different types of LISTENER RESPONSES is looking at their distribution across varying social groups. Here, I first make a binary distinction between *acknowledgements* and *other* LISTENER RESPONSE ACTIONS and then look into what these *other* Responses are for the different groups.

There are two reasons for first looking into *acknowledgements* vs all *other* Listener Responses. The primary one is largely theoretical – this split is similar to previous notions of generic and specific responses (see Bavelas et al. (2000)) and allows us to compare our findings with previous work. The secondary reason is practical – because *acknowledgements* are so common, visualising them together with the less frequent ACTION TYPES would obscure the variation between those. Hence it is more informative to zoom in on them separately, and to do so starting from the bigger picture of *acknowledgements* vs *other* in order to first establish the broad pattern.

At the broadest level possible, across all Listener-Speaker combinations, approximately 80% of all LISTENER RESPONSES are *acknowledgements*. This in turn means that only 20% of all responses fall into the *other* categories. Given that we know there is variation between male and female Listeners as well as based on the gender of the Speaker, let us break down these distributions step by step.

Figure 6.1 shows the different relative frequencies of *other* responses for male and female Listeners depending on the gender of the Speaker they are responding

FIGURE 6.1

FEMALE LISTENERS PRODUCE MORE *OTHER* RESPONSES THAN MEN, AND
SPEAKER GENDER SEEMS TO HAVE A GREATER IMPACT ON FEMALE THAN
MALE LISTENERS

to. The plot is faceted by Listener gender, with the two bars representing the frequency distribution by Speaker gender. There are three key take-away points from figure 6.1: Firstly, female Listeners produce more *other* responses than men, irrespective of Speaker gender. Secondly, Speaker gender has a stronger effect on female than on male Listeners. Specifically, and thirdly, women as Listeners produce more rather than fewer *other* responses when Listening to men.

Overall the split between *acknowledgements* and *other* actions, or to speak with Bavelas et al. (2000) 'generic' and 'specific' is fairly consistently 80\20 for female Listeners, while men do fewer 'specific' responses than women. Combined with the results from chapter 4 it seems that the male Listeners in this set of recordings respond less frequently than women, and that out of their responses a smaller proportion is something *other* than an *acknowledgement*. We will return to this difference in the first ZIP model.

One of the contributions of this chapter is to look into the diversity that the term 'specific' responses, or *other*, glosses over: Chapter 5 has shown what these actions can be, and I will now break down the different LISTENER RESPONSE ACTIONS into all categories described. Thus, the following descriptive analysis adds

Actions by Listener and Speaker gender

FIGURE 6.2
DIVERSE LISTENER RESPONSES BY LISTENER AND SPEAKER GENDERS

another layer of interactional understanding and detail to the analysis of variation in LISTENER RESPONSES.

Figure 6.2 presents each of the six OTHER ACTION TYPES in one individual facet, and each facet is further broken down by the four possible Listener and Speaker gender combinations.Percentages were calculated group-internally out of all LISTENER RESPONSES produced. To take the first bar as an example, 3.7% of all LISTENER RESPONSES women do listening to other women are *first assessments*. For ease of comparison and identification, the rounded percentage is printed above each bar. To make the ACTION TYPE labels more accessible, *self-initiated*

other-repair as described in chapter 5 is mostly referred to as *word-supply* here, and *other-initation* as *question*, given that the latter refers only to this specific sub-type (requests for information) of other-initiations.

We can make a few key observations based on this descriptive frequency overview: First of all, and most strikingly, *word supplies* are by far the least frequent across all Listener-Speaker gender combinations, making up only about 1% of all Listener Response actions within each group. *Second assessments* are similarly rare across all groups (particularly in response to male Speakers), accounting for 1.2 to 2.8% of all Listener Response Actions within each group. Furthermore, there are no stark group differences for these two actions.

The other four Action Types are somewhat more frequent, and there is more variation between the four social groups. Female Listeners do a comparatively high rate of *first assessments* (6.4%) and *surprise marks* (4.3%) when responding to male Speakers. They also do a high rate of *joint constructions* (5.1%), irrespective of the Speaker's gender. In contrast, male Listeners *ask questions* relatively more often than women, while this action is the least frequent in the all-female dyads.

6.1.1.1   Interim summary and discussion

All in all, the distributional analysis of Action Types by Speaker and Listener gender suggests that female Listeners not only produce more Listener Responses than men (see chapter 4), but at least in the data analysed here they also do more diverse actions, particularly *joint constructions* and *first assessments*. Male Listeners, by contrast, ask *questions* relatively more often. However, the goal of this PhD thesis is not to describe differences between male and female Listeners or Speakers, but rather to make a theoretical and methodological contribution that can be applied to other social variables. Let us therefore revisit the observations made above in light of the next-turn proof-procedure.

As we have seen in chapter 5 and the summary table 5.2, the definition of each Action Type is based not on the form of the response, but rather on what the Speaker has made relevant through their talk in the preceding TCU, and the orientation they display to the Response in the following TCU. Incidentally, this is precisely what Schegloff (1982) demanded from a CA perspective when he critiqued psychologists and variationists for treating all backchannels as doing the same. This means if we say 'Listener A <u>did</u> a *first assessment* in response to Speaker B' we are simultaneously saying 'Speaker B <u>provided an opportunity</u> to Listener A to produce a *first assessment*' (with reference to the preceding TCU), and 'Speaker B <u>treated</u> Listener A's Response as a *first assessment*' (with reference to the following TCU). Such an approach to coding is sensitive to the interactional dynamics and

the fact that both participants constantly co-construct the interaction as it unfolds. It also acknowledges that Listening and Speaking are inextricably intertwined and integrates this knowledge into the analysis. This integration and awareness is one of the key contributions of this chapter and the overall thesis. The next-turn proof-procedure allows us to group the Listener Response Action Types described in chapter 5 into three categories, as briefly outlined in the discussion of said chapter:

1. Those that **Listeners** can apply relatively freely and unconstrained by the Speaker's talk (*acknowledgements, questions (repair-initiation)*),

2. those that are strongly dependent on what the **Speaker** has made relevant in the TCU preceding the Listener Response (*word supplies, second assessments*), and

3. those that **both** Speaker and Listener have a potential impact on (*first assessments, surprise marks, joint constructions*).

This is important to how we interpret the results. The first group is the one in which **Listeners** are comparatively unconstrained with respect to whether and where exactly to do these actions. This is the case for most *acknowledgements* (though they can be elicited by the Speaker through rising intonation or tag questions, as well as other cues), and crucially, for *questions*. These are conceptualised as 'repair-initiation' in CA, which means the Listener treats something the Speaker has said in the preceding turn as problematic (for examples as incomplete, incorrect, inappropriate, incomprehensible). *Repair* (i.e. asking a question), as Schegloff (1982) noted, is one of the few actions that is always potentially relevant – this means, it is entirely up to the Listener whether and when to treat something as repairable. Overall, *questions*, and to a certain extent *acknowledgements*, are the Action Types that Listeners can most freely apply.

With respect to the second group, the one driven mainly by the **Speaker**'s actions, recall that *word supplies* and *second assessments* were extremely rare across all groups (though there was some between group variation). Both *word supplies* and *second assessments* are highly contingent on opportunities provided by the Speaker in the preceding talk. If the Speaker does not signal word-search trouble, the Listener has no reason whatsoever to provide a *word supply*. Attempts to do so would more likely be interpreted as a claim to the floor. Similarly, a *second assessment* by definition follows a first assessment, so in order for the Listener to produce a *second assessment*, the Speaker must have done a first assessment in the preceding TCU. Hence, if we say a Listener does few *word supplies*, we are simultaneously saying the Speaker they are responding to rarely makes *word*

supplies relevant. In fact the frequency of *word supplies* and *second assessments* says more about how often the Speaker <u>makes them relevant</u> than about how many the Listener <u>chooses to produce</u>. This is the case because it is almost obligatory to 'help' the Speaker by suggesting a possible word, if word-search trouble is being signalled.

The third group of Action Types discussed here is characterised by equal involvement of **both** Speaker and Listener. It comprises of *first assessments, surprise marks*, and *joint constructions*. For all of them, opportunities to produce them are constrained by how both participants jointly co-ordinate and organise their interaction. A variety of aspects in the Speaker's talk can be assessable or surprising from the Speaker's perspective, and/or taken up as such by the Listener. Which of these things are picked up is then negotiated in the interaction. This applies similarly to *joint constructions*. Listener involvement can be considered slightly higher here than for the other two ACTION TYPES just discussed. Thus, when talking about the frequency of these three ACTION TYPES, we are making statements about the Speaker's and the Listener's behaviour *together*.

This three-way distinction based on which party has how much impact on the specific Action being done will be used in the regression models in the next section.

6.1.2   Regression models

We can *describe* what is happening in the data at hand based on a distributional analysis, but we must be very careful not to generalise from the descriptive statistics for three reasons: First of all, group-averages might be driven by individual Listeners or dyads, and descriptive statistics cannot factor this in. Second, descriptive statistics become increasingly convoluted and difficult to interpret the more groups we add – here, I was able to include Speaker and Listener gender in the analysis, but more would have become too complex to present in a descriptive overview. And a third related issue is that descriptive statistics do not readily allow us to portray interactions between different factors.

These concerns can be addressed using inferential statistics, specifically Zero Inflated Poisson (ZIP) regression models, which were introduced in the methodology, and in chapter 4. For the present chapter, I have fit these models with `brms` (Bürkner 2018) as Bayesian models, because they allow for the inclusion of random slopes and intercepts (on the advantages of random intercepts see also Drager and Hay (2012)). Some Listeners in some interactions do not do all of the different actions, leading to cells with zeros. In order to model those, we need the Zero-inflated part of the ZIP model. Because the vast majority of this variability is

due to random effects of Listener and Dyad, the analysis focuses on the output of the count model.

The models are constructed as follows: the outcome variable is the number of actions of any given type any Listener does in any given Dyad, and the fixed effects are ACTION TYPE interacting with Speaker gender and Listener gender. Because some of the effects are driven by individual Listeners or Dyads, random slopes and intercepts by ACTION TYPE were added for Listener and Dyad respectively. In order to account for the variable length of conversations and amount of talk done by each participant (see appendix I), the log number of Speaker words in each conversation is used as an offset variable. The offset factors in that the Listener has more space to produce Responses, the longer the Speaker talks. This is analogous to the finding that number of Listener Responses increases with Speaker turn length presented in chapter 4.

Coming from a less interaction-focused perspective, it might seem intuitive to include turn length as a predictor for how often which action is done, rather than just using length of Speaker-talk in the conversation as an offset. However, from an interactional perspective, turn length is irrelevant to *which* actions get done. Actions are not (made) relevant based on the *amount* of the preceding talk, but rather based on the *actions* done in and made relevant by it. In some cases turn length might mask a different process, leading us to think that longer turns elicit more diverse Listener Response actions. For example, if we analyse narrative vs non-narrative talk, Listeners produce more 'other' Action Types in response to narratives, and which specific Action Types they do depends on which part of the narrative they are responding to. Narratives tend to be longer turns than other parts of the interaction, but again, the aspect of talk that makes a certain Action Type relevant in response is not the length of the turn but rather what the Speaker *does* in that turn. I ran one model including turn length, and it did not improve the model fit, nor did it make a difference relative to the models I present.

In summary, the model predicts how many Actions of any given type a male/female Listener is likely to do in response to a male/female Speaker in any given conversation, based on the amount of talking the other person does in said conversation. The model formula developed based on this qualitative understanding of the data is as follows:

> brm(number of responses ~Action Type * (Listener gender + Speaker gender) + (1 | Action Type + Listener) + (1| Action Type + Dyad) + offset = log(number of Speaker words in conversation), family = zero_inflated_poisson, prior = priors)

Action Type as the variable has different variants in the three models presented below: In the first model, the variants are *other* and *acknowledgement*, in the second model all seven individual Action Types are included as variants, and in the third model the actions are recoded into three types, based on who has the highest impact on them getting done: the *Listener*, the *Speaker*, or *both* (see the interim discussion following the descriptive statistics).

Priors were specified based on the output of the `get_prior()` function, which is part of the `brms` package. The steps outlined here briefly are described in more detail in the `brms` package documentation (see also Bürkner (2017), Bürkner (2018), and Carpenter et al. (2017)). Thus, for all three models, mildly informative priors that are based on the observed distributions were chosen.

Following the structure of the previous section, I will first present the ZIP model where TYPE OF ACTION is a binary variable with the variants *acknowledgements* and *other actions*, then present the ZIP model that predicts counts for all seven Action Types, and finally discuss a ZIP model for three groups of Actions, based on how much impact the Speaker or Listener respectively have on the action being done.

6.1.2.1   ZIP model for binary distinction

Bayesian statistical modelling is different from Null Hypothesis Testing in that its results are probability distributions, called posterior distributions. Thus, in the regression tables presented below, the Estimate represents the mean of the posterior distribution. We can interpret this as the most *likely* outcome. The 2.5 and 97.5% credible intervals (CIs) represent the spread of the posterior distribution, and thus the *confidence* in our predicted estimate – if the CIs are narrow, the model is fairly confident in the estimate, while if the CIs are wide, there is a lot of uncertainty about the results.

Reading Bayesian regression output therefore differs slightly from reading non-Bayesian output: in non-Bayesian models the estimate gives us all the information we need, but in a Bayesian model we need the estimate *together* with the credible intervals, because these intervals tell us how 'sure' the model is about its prediction based on the data. This also explains why we are not concerned about statistical power in the same way we would be in a non-Bayesian framework; the model predicts not only the most likely estimate, but also informs us about the 'certainty' of that prediction, or in other words about the likelihood of the prediction being correct.

How this works will become evident as we step through the model output in table 6.1. The model takes the number of *acknowledgements* done by female Listeners towards female Speakers as the baseline, hence the regression table

reports how this number changes when the Action Type is *other*, or when the Listener or Speaker respectively is male. There are two notable effects, one that we can be very certain about, and one with a reasonable high level of certainty. In the following, the clearly positive or negative effects are marked up in bold in the regression table, and the trends with some level of uncertainty are underlined.

What the model predicts with a high level of certainty is that there are fewer *other* Listener Responses than *acknowledgements*. The estimate for this is -1.578 (in log odds), and 95% of the possible scenarios our Bayesian model has come up with fall between -2.005 and -1.173. Thus, we can be very certain that the effect is negative, and there is a relatively small degree of uncertainty about the size of the effect. This results reflects the 80\20 split reported in the descriptive statistics above.

	Estimate	Est.Error	CI 2.5%	CI 97.5%
(Intercept)	-3.237	0.147	-3.522	-2.945
Action Type (other)	**-1.578**	0.210	**-2.005**	**-1.173**
Listener gender (male)	<u>-0.331</u>	0.242	<u>-0.812</u>	<u>0.142</u>
Speaker gender (male)	0.005	0.101	-0.195	0.197
Action Type (other) x male Listener	<u>-0.224</u>	0.369	<u>-0.963</u>	<u>0.465</u>
Action Type (other) x male Speaker	0.025	0.151	-0.259	0.334

TABLE 6.1
FIXED EFFECTS OF THE ZIP MODEL PREDICTING HOW MANY *OTHER* ACTIONS AND *ACKNOWLEDGEMENTS* MALE OR FEMALE LISTENERS DO.

The second trend that we can be reasonably certain about is that male Listeners produce fewer responses than female Listeners: The estimate is -0.331, and the 95% Credible Intervals (CIs) are -0.812 and 0.142. This means that a small part of the posterior distribution is positive, but the bulk of the likely observed values is in the negative space. Hence, we can be fairly confident in saying that male Listeners produce fewer Listener Responses than female Listeners, but there is a lot of individual variability.

By contrast, Speaker gender does not have a clear overall influence on the number of Actions done: the Estimate (i.e. the change in the intercept) is very close to 0 at 0.005, and the 95% CIs are symmetrical around 0, ranging from -0.195 to 0.197. This means we cannot claim that Speaker gender overall has an impact on how many *acknowledgements* and how many *other* actions get done.

Looking at the interaction effects, the same overall pattern holds, though the effect of Listener gender is smaller and has a high degree of uncertainty: male

Listeners are less likely (estimate = -0.224, CIs -0.963 to 0.465) to produce *other* Listener Responses than female Listeners. There is a lot of uncertainty about the exact effect size, including about the direction of the effect, with the posterior distribution ranging from -0.963 to positive 0.465. Hence, it is possible that the variation observed is due to individual variability rather than Listener gender. The interaction between ACTION TYPE and Speaker gender mirrors the one seen for Speaker gender only, but with even more uncertainty (i.e. wider CIs, as seen in the interaction between Listener gender and ACTION TYPE).

Overall, this first model shows that there are fewer *other* actions being done (negative estimate for 'other', and CIs also in the negative space), and that Listener gender impacts on their number, with male Listeners producing the fewest *other* responses when responding to male Speakers (Interaction between 'Action type (other)' and 'male Listener', though this effect is far smaller and the CIs wider than for the effect of 'Action Type (other)' alone). This reflects the pattern seen in the descriptive statistics, as well as the overall frequency results presented in chapter 4.

6.1.2.2 ZIP model for all action types

This second model is structured just like the above, but the dependent variable, ACTION TYPES, has seven variants: *acknowledgements, first assessments, second assessments, surprise marks, joint constructions, questions*, and *self-initiated other-repair (word-supplies)*. Again, all responses of each ACTION TYPE were counted up per Listener per conversation, as were the number of Speaker words in each conversation. The model output is summarised in table 6.2. Based on previous results and the categorization of ACTION TYPES according to which party has how much influence on whether this action gets done, I hypothesise the following:

1. We expect Listener gender to have a greater overall impact on the number of responses produced than Speaker gender, following the results of the first model presented above and the model on overall frequency presented in chapter 4,

2. We expect nuances in the interactions between ACTION TYPE and Listener and Speaker gender:

    (a) for *acknowledgements* and *questions (other-initiation)*, Listener gender has a greater effect than Speaker gender;

    (b) for *word-supplies* and *second assessments*, Speaker gender has a greater effect than Listener gender;

(c) for *joint utterances, first assessments*, and *surprise marks*, Listener and Speaker gender have similarly strong effects.

	Estimate	Est. Error	CI 2.5%	CI 97.5%
Intercept	-3.235	0.156	-3.540	-2.936
Effect of Action Type				
first assessment	-3.080	0.240	-3.564	-2.633
joint construction	-2.899	0.250	-3.410	-2.422
surprise mark	-3.802	0.495	-4.849	-2.854
question	-4.055	0.557	-5.212	-2.976
second assessment	-3.441	0.288	-4.050	-2.915
self-initiated other-repair	-4.821	0.491	-5.875	-3.976
Effect of Listener and Speaker gender				
Listener gender (male)	<u>-0.329</u>	0.268	<u>-0.868</u>	<u>0.218</u>
Speaker gender (male)	-0.001	0.101	-0.201	0.188
Interaction Action Type x Listener gender (male)				
first assessment x male Listener	**-0.687**	0.384	**-1.481**	**0.062**
joint construction x male Listener	**-0.948**	0.462	**-1.908**	**-0.037**
surprise mark x male Listener	-0.110	0.800	-1.673	1.514
question x male Listener	-0.056	0.936	-1.959	1.688
second assessment x male Listener	-0.195	0.446	-1.075	0.726
self-initiated other-repair x male Listener	-0.417	0.795	-2.085	1.033
Interaction Action Type x Speaker gender (male)				
first assessment x male Speaker	<u>0.323</u>	0.273	<u>-0.205</u>	<u>0.855</u>
joint construction x male Speaker	-0.038	0.192	-0.397	0.345
surprise mark x male Speaker	-0.017	0.353	-0.729	0.687
question x male Speaker	-0.011	0.363	-0.718	0.716
second assessment x male Speaker	**-0.739**	0.316	**-1.347**	**-0.104**
self-initiated other-repair x male Speaker	<u>0.434</u>	0.444	<u>-0.363</u>	<u>1.377</u>

TABLE 6.2

FIXED EFFECTS OF THE ZIP MODEL PREDICTING HOW MANY ACTIONS LISTENERS DO BASED ON LISTENER AND SPEAKER GENDER FOR ALL SEVEN ACTION TYPES.

The model uses *acknowledgements* as the baseline to compare the number of occurrence of the other actions to. Thus, it is unsurprising that the main effect of all

individual Action Types is negative, given that *acknowledgements* are by far the most frequent Action Type. These effects are given in the first block of results in table 6.2.[3] It is worth noting that even though the effect relative to *acknowledgements* is clearly negative for all other Action Types, the certainty of the estimates varies. The stronger the negative effect, the lower the certainty (i.e. the wider the CIs). Specifically, certainty is relatively high for the estimates of *first assessments, joint constructions*, and *second assessments*, but only half as high for *surprise marks, questions*, and *self-initiated other-repair (word supplies)*.

With respect to Listener and Speaker gender, there is a clear trend for male Listeners to produce fewer responses than female ones. However, a small part of the posterior distribution is in the positive space, rendering this result a trend rather than a clear effect. Speaker gender again has no marked effect, with an estimate close to 0, and the 2.5% and 97.5% CIs almost symmetrical around the mean.

The interactions between the individual Action Types and Listener gender (third block of results) and Speaker gender (fourth block) respectively are more interesting, and this is where we return to the predictions listed above. Let us begin with the categories that show no or only very small effects: Speaker gender has no clear effect on *joint constructions*, neither Listener nor Speaker gender have a clear effect on the frequency of *surprise marks* or *questions*, and Listener gender has no clear effect on *second assessments*.

When it comes to *self-initiated other-repair (word supplies)*, there seems to be a slight trend for male Listeners to produce fewer *word supplies* than women, but the posterior distribution is extremely wide, ranging from -2 to +1. This indicates a large amount of uncertainty in the estimate. By contrast, the effect of Speaker gender on the number of *word supplies* done in any given interaction shows a positive estimate at a reasonable degree of certainty. The 95% CIs range from -0.363 to 1.377, suggesting that the bulk of the posterior distribution is in the positive space. Hence, we can be reasonably sure in claiming that Listeners are more likely to produce *word supplies* when listening to a male than to a female Speaker, even when we take into account variability due to characteristics of the individual and the dyad.

A possible explanation of this pattern from a language and power perspective is that Listeners tend to behave in a more facilitative manner towards male than female Speakers. From a 'difference' perspective we could also interpret this as 'male Speakers make *word supply* relevant more often than women'. This

---

[3]In this and the next model, the main effects of ACTION TYPE on its own are not printed in bold even if the results show clear effects, because this is strongly expected based on the descriptive statistics. I decided to visually emphasise only the more interesting interaction effects that cannot be gleaned from the descriptive statistics.

second interpretation is in juxtaposition with the patterns observed with respect to frequency, where male Speakers hold onto the floor more strongly than female Speakers by talking for longer without having received a LISTENER RESPONSE, and 'tolerating' receiving fewer responses than female Speakers. Here, they are making themselves more vulnerable by producing opportunities for the Listener to come in and supply a word (or even a phrase, and then take over the floor). However, whether men really offer more opportunities for Listeners to do *word supplies*, or whether Listeners produce *word supplies* more readily towards male than female Speakers is an empirical question that warrants further study and necessitates full coding of all actions in all conversations.

Similarly to *word supplies*, *second assessments* show an effect based on Speaker but not Listener gender. When responding to a male Speaker, Listeners produce fewer *second assessments* than when Listening to a woman (estimate = -0.739, 95% CIs -1.347 to -0.104). This could again be interpreted in a number of ways relative to language, power, and gender. What I outline here are hypotheses which would need to be tested by – as just stated above – coding all the conversations for all the actions, or an in-depth qualitative analysis focusing on assessments.

The first interpretation, and the one I build on later, is that male Speakers provide fewer opportunities to produce *second assessments*. This assumes they do fewer (first) assessments than female Speakers. Assessments can be closing-implicative, hence doing an assessment as a Speaker involves an elevated risk of losing the floor (see for example Goodwin and Goodwin (1987b), as well as the third possible interpretation). An analysis focused on language and power might argue that male Speakers avoid risking the floor by avoiding assessments.

A second possibility is that male Speakers do assessments just as frequently as women, but Listeners produce *second assessments* less often towards male Speakers. This could be seen as orienting to the Speaker's right to keep the floor, and suggest that Listeners are more supportive and facilitative towards male than female Speakers. Such an interpretation would corroborate similar notions of language and power, with male Speakers treated as having more right to the floor.

On the other hand, and this is the third possibility, male and female Speakers might be doing assessments at the same rate, but when Listeners respond to male Speakers with a *second assessment* this leads to a change in the floor more often. These *second assessments* would not show up in the 'Listener Response' count because they are then 'new-Speaker talk'. If this were the case, Listeners could be seen as *less* facilitative towards men. It is left to future work to investigate which of these is the case, and how male and female Listeners and Speakers are actually performing gender at the local level of the interaction. What I hope this brief discussion of possible interpretations has demonstrated is that it would be

hasty to jump from the effects observed here to claims about gendered behaviours reflecting innate differences or established power structures.

Returning to the model summary shown in table 6.2, there is a negative effect of Listener gender (male) on the number of *first assessments* and *joint constructions* done in any given interaction, but the CIs around both are comparatively wide. Thus, even though we can be fairly certain about the direction of the effect, there is a large degree of uncertainty about its size, suggesting that a fair amount of the variability observed is due to individual characteristics or random factors to do with the dynamics of the dyad. This also applies to the slight positive trend observed in the interaction between the Speaker (person being Listened to) being male, and the Action Type being a *first assessment*: it looks as though male Speakers are slightly more likely to receive a *first assessment*, but part of the posterior distribution is in the negative space, rendering this a trend rather than an effect.

Summarising the effects observed in this model, we can review the likelihood of the hypotheses presented above as follows:

1. Listener gender indeed has a stronger impact than Speaker gender overall – in fact Speaker gender overall is very close to having no effect at all.

2. There are nuances with respect to the interaction effects based on Action Type and Speaker and Listener gender:

   (a) for *acknowledgements* and *questions (other-initiation)*, **Listener** gender is indeed more important than Speaker gender, however, the effect on *other-initiations* is very small in both cases (though it does seem to be slightly stronger for Listener gender, with the posterior distribution being slightly heavier in the negative space rather than symmetrical around 0);

   (b) for *word-supplies* and *second assessments*, the effect of **Speaker** gender is indeed more important than that of Listener gender.

   (c) For all others, **both** Listener and Speaker gender have similarly strong (or rather similarly weak) effects.

Particularly the finding with respect to *word-supplies* and *second assessments* is striking considering that Listener gender is overall more important in all models presented, while Speaker gender has only a very small effect, if at all. These results call for a third model based on Main Actor as a variable, and *Speaker, Listener* or *both* as variants. This model is presented in the next section.

6.1.2.3 ZIP model for three types of Listener Responses

When the individual actions are re-coded into three groups – those that the Listener interactionally has the most impact on, those that the Speaker has the most impact on, and those that both parties contribute to equally (see the interim summary and discussion, section 6.1.1.1) – we see an interaction effect between TYPE OF RESPONSE and Speaker or Listener gender respectively: Male Listeners are more likely than female Listeners to produce those actions that the Listener has more impact on, and Listeners are less likely to produce those responses that the Speaker has more impact on when responding to a male Speakers (see the bottom two tiers of table 6.3).

	Estimate	Est. Error	CI 2.5%	CI 97.5%
Intercept	-5.227	0.218	-5.676	-4.799
Effect of Main Actor				
Listener	2.020	0.218	1.588	2.446
Speaker	-1.151	0.204	-1.556	-0.751
Effect of Listener and Speaker gender				
Listener gender (male)	**-0.947**	0.387	**-1.700**	**-0.188**
Speaker gender (male)	0.138	0.106	-0.071	0.342
Interaction Main Actor x Listener gender (male)				
Listener x male Listener	**0.621**	0.376	**-0.121**	**1.374**
Speaker x male Listener	<u>0.421</u>	0.321	<u>-0.221</u>	<u>1.050</u>
Interaction Main Actor x Speaker gender (male)				
Listener x male Speaker	<u>-0.138</u>	0.134	<u>-0.405</u>	<u>0.123</u>
Speaker x male Speaker	**-0.524**	0.230	**-0.986**	**-0.074**

TABLE 6.3
FIXED EFFECTS OF THE ZIP MODEL PREDICTING HOW MANY ACTIONS
LISTENERS DO BASED ON WHICH PARTY HAS HOW MUCH IMPACT ON THE
ACTION FROM A QUALITATIVE PERSPECTIVE.

The model takes Actions both participants contribute to equally as a baseline and compares the other two groups to this. The first tier, 'Effect of Main Actor', simply confirms that there are more of the responses that are mainly affected by the Listener, and fewer that are mainly affected by the Speaker. Neither of these effects is surprising, given that the Listener-driven ones include *acknowledgements*, by far the most frequent Action, while the Speaker-driven ones include the two least frequent ACTION TYPES. The second tier summarising the effects of Listener and Speaker gender shows that overall Listener gender matters more than Speaker gender, and men do a smaller number of responses overall than women. This is in line with all other models presented so far.

The most important sections of the regression table are the two at the bottom that show the interactions between Main Actor and Listener and Speaker gender respectively. Together with Main Actor being *Listener*, Listener gender has a clear effect, with men producing more *acknowledgements* and *questions* than women. Given the overall strong effect of Listener gender, we also note that male Listeners are more likely to do those types of Actions that the Speaker has a stronger impact on (estimate = 0.421, 95% CIs -0.221 to 1.050).

Conversely, the number of Speaker-driven Listener Responses is indeed more strongly predicted by Speaker gender than by Listener gender (estimate = -0.524, 95% CIs -0.986 to -0.074). This suggests that, even when accounting for individual and dyad-related random variability, Listeners produce fewer of these responses when listening to male than female Speakers. It is not surprising that Listener gender still has a small effect even for those Speaker-centred actions, given its strong individual effect across all models and conditions.

Overall this third model where the Action Types are coded based on which party has how much impact on them shows that the Speaker-driven actions interact with Speaker gender (getting done less often when the Speaker is a man), and the Listener-driven actions are more frequent when the Listener is a man. In both cases, the gender of the other person has a small effect with a large degree of uncertainty, including about the direction of the effect.

6.1.3   Summary of all results in this chapter

Overall, the descriptive analysis has shown that for female Listeners, *other* Listener Responses make up a higher percentage than for male Listeners, and that both men and women as Listeners increase their frequency of *other* responses in mixed-gender dyads relative to same-gender dyads (figure 6.1, thought the difference is much more pronounced for female than male Listeners). When zooming in on the individual Action Types, we have seen that women produce more *joint utterances* and *first assessments* than male Listeners, while men as Listeners ask more questions. We have seen further variation based on Speaker gender. Given that there is individual variation related to both Speakers and Listeners, some of these effects are likely driven by individual Listeners or Dyads.

The ZIP models take into account these random effects of individual Listener and Dyad dynamics. They show that overall, Listener gender has a far stronger effect than Speaker gender, with male Listeners producing fewer 'other' responses overall. When it comes to the interaction between Speaker or Listener gender and the individual Action Types, men produce fewer *first assessments* and *joint constructions* but approximately the same number of *questions* than women (see

165

central tier of table 6.2). With respect to Speaker gender, Listeners of either gender produce more *first assessments* and *questions*, and fewer *second assessments* when responding to male Speakers. When the different ACTION TYPES are aggregated into the ones that interactionally mainly the *Speaker*, the *Listener*, or *both* participants have an impact on, this effect is reflected in the posterior distributions: even though Listener gender is a much stronger predictor overall, Speaker gender is the more important predictor for the number of 'Speaker-driven' Listener responses done by Listeners of either gender.

6.2   Discussion

As stated in the introduction to this chapter, the analysis just presented makes three key contributions towards integrating interactional and variationist methods:

1. It applies the taxonomy developed in the previous chapter to the data, thereby firmly rooting the analysis of variation in which ACTION TYPES get done in the interaction.

2. It demonstrates that using the next-turn proof-procedure uncovers systematic variation in the data and impacts on how we need to interpret the results:

   (a) *Acknowledgements* and *questions (other-initiation)* are mainly driven by the **Listener**, while

   (b) *word-supplies* and *second assessments* are mainly driven by the **Speaker**, and

   (c) the others are impacted by **both** to a similar extent.

3. Finally, Bayesian ZIP models are presented as an inferential statistical method that allows us to model count data including zeros or empty cells and to include complex fixed and random effect structures without the usual convergence issues.

This is the third step in closely integrating an awareness of interactional structure in the quantitative analysis of variation. We begin to see that on all levels, the structure of the interaction and in particular the actions or linguistic production of the Speaker (ie the 'other' participant) are key in defining, delimiting, quantifying, and interpreting the variable and variants. The final discussion (chapter 8) focuses on the impact and implications of the next-turn proof-procedure, and I have argued for the need for inferential models like the one presented here in the interim discussion in this present chapter. Therefore, I will focus on contextualising the contributions made in the thesis overall, before moving on to the final analysis

chapter that ties the analysis back to the exact lexical and prosodic realisation of the Listener Responses.

As discussed in the context of the frequency analysis in chapter 4, in variationist sociolinguistics the envelope of variation is generally taken to reside in the individual producing the variable under analysis. This assumption has been transferred to the conceptualisation of Listener Responses as a variable, even in studies where scholars have then looked at the frequency of different *types* of Listener Responses (see for example Namba (2011), Stubbe (1998), Tolins and Fox Tree (2014), and Tottie (1991)).

The previous chapter (5) developing the taxonomy of Listener Responses has demonstrated how interaction is constantly co-constructed by both participants. This idea is not usually applied and translated into methodological decisions and analytical considerations in sociolinguistics. An obvious exception is Speech or Communication Accommodation Theory (Giles 1973), which posits that speakers design their talk with respect to their interlocutor. However, convergence or divergence are generally theorised to be based on social distance and relationship management or attitudes to the other, and the interlocutor's speech patterns are taken as a given. Llamas et al. (2009) for example compare participants' speech with an English, a Scottish, and a 'neutral' interlocutor and find different patterns of convergence and divergence for different variables. However, they do not take the actual production of these interlocutors into account – this is a common approach in CAT, coming from a social psychology perspective and focussing more on attitudes and social distance than interaction.

Thus, even in the theoretical approach that does take the interlocutor to be *the* crucial conditioning factor of language variation, the sociolinguistic interview is used to elicit speech from the informant, and the analysis is focused on the interviewee's talk but not on the *interaction* said talk is couched in. Labov (2013) mentions this as a challenge, particularly when analysing the development of narratives, but this awareness is rarely taken up in studies focussed on language variation – and if it is, then only to highlight individual instances, rather than as a coding scheme or even an overall analytic orientation (see for example Buchstaller (2008), Eckert (2009), and Eckert (2011)). In a study of quotatives, Buchstaller (2011) shows that participants use different quotatives in narratives and other talk, but this observation is only a small part of a much bigger analytic project.

So why and how does it matter to code for and take into account these different interactional roles, and the different contributions Speaker and Listener make to any Listener Response doing the Action it does? Because it is not only reductive, but in fact misleading to ascribe any individual action to any single participant in the interaction a priori.

The overall frequency of Listener Responses has been shown to strongly depend on the length of the turn that is being responded to, rather than talk produced by the person doing Listening (see chapter 4). Thus, the conditioning factors of variation are situated in the interaction rather than the individual. The same applies to variation in the frequency of Action Types: The primary structural conditioning factor is the talk (or rather action) produced by the Speaker in the preceding TCU. Some Speaker-actions strongly constrain which actions can or even need to be done next, while others do not impose such constraints. This pattern of interactional constraints is visible in the posterior distributions of the second and particularly the third model presented above. The strength of the effects of Speaker and Listener gender respectively reflect how strongly the preceding Speaker-talk constrains the possible next action(s).

It is an important contribution to variationist sociolinguistics to tease this apart, because it demonstrates that we must not look at the actions done solely as 'things the Listener decided to do'. Instead, we need to differentiate based on how much Speaker and Listener interactionally contribute to each action and take this into account when quantifying as well as interpreting the results.

One potential concern might be that the fact that the Listener is not the only person accountable for Listener Response Actions could undermine the validity of the Action Type coding. This is not the case, though: As discussed in the section on the next-turn proof-procedure, the fact that the conversation usually runs smoothly and one party does not challenge the other's interpretation of their actions suggests that the categories observed and described are valid, particularly because their definitions are based on the participants' behaviours.[4]

In the analysis so far, we have seen that it is crucial to take into account the Speaker's linguistic production when operationalising overall Listener Response frequency (chapter 4), and that depending on the Action Type, the Speaker, here seen through the social variable gender, has a greater effect on their occurrence than the Listener. This points to the interaction as the locus of important conditioning factors constraints on variation for Listener Responses on several levels. In the next and final analysis chapter, we will see that the Speaker's linguistic production also impacts on the exact form of the individual responses at the levels of lexis and prosody. I will develop the idea of the preceding Speaker-talk as a crucial structural conditioning factor on the realisation of the individual utterances.

---

[4]To reiterate: If a Speaker treats a Listener Response as doing X, and the Listener does not problematise this by initiating repair, this indicates that the Speaker's next action displays an understanding close enough to the Listener's 'intended' action. When the Listener does *not* contest the way in which the Speaker orients to their response, this interpretation can be considered 'correct', close enough to the 'intended action', or at least an 'acceptable' understanding of the preceding action. Accordingly, it is reasonably safe to assume that Listener has done the action that the Speaker orients to in their next turn.

Chapter 7

How are the Actions done? The Form-Function Link for Action Types

> An unguarded glance, a momentary change in tone of voice, an ecological position taken or not taken, can drench a talk with judgemental significance.
>
> *Erving Goffman* (1982: 33)

This final analysis chapter is based on the observation that Speakers generally 'understand' and correctly orient to the Action a Listener does in any given Response, even though some Action Types might be done using the same lexical material.[1] Assuming that one 'form' is linked to one function to minimise ambiguity, this suggests 'form' is underspecified based on lexical material alone. 'Lexical material' here is used to refer to both words and vocalisations like 'mhm' and 'uh-huh' and was chosen as a term because 'words' might suggest vocalisations which are limited to what one might find in a lexicon.

The four analysis chapters have all relied on the level of *discourse-organisation* as the level of structure at which Listener Responses and the individual Action Types are situated. When analysing the relationship between the *realisations* (variants) of these individual ACTION TYPES (variables) at the vocalised level, this has three theoretical and methodological key consequences, which I state here briefly and then discuss:

---

[1]As shown in chapter 5, function, or Action Type, is defined based on sequential structure (what happens in the preceding and following TCUs) rather than on form. This allows us to explore variation in the realisation of each Action Type.

1. Differentiating between Action Types based on their sequential structures and interactional impact (rather than defining them based on their form) is what allows us to analyse variation in *how* these Actions are done.

2. The preceding and following TCU (i.e. talk produced by the Speaker) constitute the relevant context of occurrence and are hypothesised to impact on the exact realisation of the variant.

3. 'Form' needs to be described on all the levels relevant in that moment in the interaction. It is based on a complex interplay of features beyond lexical choice. For practical reasons, I focus on the vocal modality here and demonstrate that at least in some cases prosody is an integral part of the form.

We can directly compare the definition of discourse-organisation and phonological variables as well as variation in the realisation (and perception) of the variants. This applies with respect to the preceding and following talk as conditioning factors on the precise lexical, prosodic, and morphosyntactic realisation of the Listener Response, as well as with respect to the form-function link: 'form' is specified on a number of levels of linguistic structure which stand in a complex interplay with each other. I will use the phonological contrast between the bilabial plosives [p] and [b] as an example.[2] Table 7.1 shows how the levels of structure relevant to a phonological variable apply and can be transferred to a discourse-organisational one like Listener Response Action Types. Overall, phonological variables are defined on the level of phonology (sometimes in interaction with other traditionally 'linguistic' levels like morphosyntax or semantics) and their variants are described in detail on the level of phonetics. Discourse-organisational variables like the individual Listener Response Actions, on the other hand, are defined based on sequential structure. The scope of the present investigation is to describe the individual variants on the lexical and prosodic level.

7.0.1 Implications of a structure-based definition

Stepping through table 7.1 from top to bottom, the first parallel is drawn at the level of defining and identifying instances of the variable. It might seem entirely intuitive that bilabial plosives are defined based on their position in a word, or when that is ambiguous, for example because the word forms a minimal pair,

---

[2]I would like to thank the Glasgow University Laboratory of Phonetics audience from January 2019, in particular Rachel Smith, for their comments on an earlier version of this analysis, and for pointing out the parallel between variable realisations of bilabial plosives and Action Types as I have presented them here.

	Phonological	Discourse-organisational
DEFINITION	**[p]-[b]**	**ACTION TYPES**
	position in lexical item	position in discourse
FACTORS INFLUENCING PRODUCTION AND PERCEPTION OF FORM		
Coarticulation	preceding phoneme	preceding TCU
Assimilation	following phoneme	reaction to following TCU, if started in overlap
Other	length of preceding vowel	position relative to TCU: overlap, in the clear, towards the edges
	Position in word (initial, medial, final)	Position in longer project
	voice quality	
	word and sequence stress	
	speech rate	
CORRELATES OF FORM		
Core	Voice onset time (VOT)	lexical material
Other	aspiration	lexical frequency
	voicing	amplitude
	closure and release	pitch and pitch movement

TABLE 7.1

INFLUENCE OF PHONOLOGICAL STRUCTURE ON PHONOLOGICAL VARIABLES AND SEQUENTIAL STRUCTURE ON DISCOURSE-ORGANISATIONAL VARIABLES.

based on semantic and morphosyntactic information. It might seem less intuitive that exactly the same applies to Action Types: they are identified based on their position in the discourse and impact on its development. This structure-based definition, which we are familiar with from phonological variables, allows us to then analyse variation in the actual realisation (or form) of the individual Action Types.

The second part of table 7.1 can be thought of as representing the envelope of variation. It refers to different types of factors influencing the production and perception of the individual variants, particularly those that go beyond the correlates of form listed above. This means the notions of 'coarticulation' or even 'assimilation' we are familiar with from phonological variables can be applied to discourse-organisational variables like Listener Response Action Types. For phonemes, 'coarticulation' is based on the *sounds* surrounding the variable or

more general characteristics of the talk produced by the person producing the sound under investigation. To give an example: if I wanted to study DaisyRae's realisations of [p] when talking to Rose, I would take into consideration which phonemes DaisyRae produces before and after the variant, what her speech rate is, her voice quality, where applicable the length of the vowel she produces before the variable I am interested in, etc. Crucially, in all these cases I am interested in DaisyRae's production, not in Rose's.[3]

For the discourse-organisational variable the immediately relevant surrounding context are the *preceding and following talk* (as well as the broader context of how the interaction develops) and hence the linguistic production of the Speaker (i.e. the other person in the interaction). Let us take DaisyRae and Rose's conversation as an example again. If I want to describe the exact realisation of an ACKNOWLEDGEMENT that DaisyRae does, I need to take into account the preceding talk produced by Rose. If Rose's preceding TCU has a high amplitude, DaisyRae's ACKNOWLEDGEMENT can also have a relatively high amplitude, as long as it is lower than Rose's. If Rose has been speaking quietly, DaisyRae's response needs to be very quiet to still be prosodically marked as an ACKNOWLEDGEMENT. With respect to the lexical level, if Rose signals a word search, DaisyRae needs to provide the word she thinks Rose is looking for, not just any odd word, and in a joint utterance DaisyRae would most likely recycle at least some of the lexical material Rose has used in the preceding TCU. This explains why 'form' on the lexical as well as the phonetic level needs to be considered relative to the preceding talk uttered by the *Speaker* rather than the Listener.

Similarly to the observation that prosody is an integral part of the form, scholars developing a grammar (Ochs et al. 1996; Thompson et al. 2015), a prosody (Szczepek Reed 2007), and a phonetics of talk-in-interaction (Local et al. 1986; Ogden 2006; Ogden 2012; Gorisch et al. 2012) have noted that each turn or TCU is formatted with close attention to the design of the preceding talk. Nilsson (2015) is one of the few scholars who has integrated this knowledge into an analysis of language variation, specifically dialect accommodation in Swedish. She shows that when participants format-tie, they are more likely to use the dialect features their interlocutor has just used in the preceding TCU, even if those variants are rare in the rest of the format-tying participant's speech. Furthermore, Raymond (2018) looks at participants' orientations to phonological, morphosyntactic, and lexical variants associated with different varieties of Spanish, and links this to variationist models of dialect contact, arguing that Conversation Analysis allows us to understand how participants use dialectal variation in situ and can help

---

[3]I might consider Rose's realisations of [p] separately, and then analyse whether DaisyRae and Rose accommodate to each other, but that is a different analysis.

explain both language change and stability in contact situations. Thus, integrating interactional knowledge and methods into the variationist paradigm allows us to better understand variation at all levels of linguistic structure.

The third parallel is at the level of describing features that make the individual variants recognisable and distinctive. This is represented in the third tier of the comparison table, 'Correlates of form'. For both the phonemic contrast and the Action Types, 'form' consists of a complex interplay of features on different levels, not just the core contrast that in many cases characterises the difference very well. VOT is taken to be the core contrast for bilabial plosives, while for Listener Responses this tends to be the lexical material. However, phoneticians have noted that in fact the production and perception of the individual variants are characterised by not only VOT but also by aspiration, voicing, closure, and release (for example Raphael (1972), Williams (1977), Macken and Barton (1980), and Flege and Brown (1982); for an excellent overview see Ogden (2017)), and that these cues in concert acoustically distinguish one set of variants from the other.

The same is true for the individual Action Types, as we will see in this chapter: There are several relevant characteristics on the lexical and prosodic levels that contribute to meaningfully distinctive forms for each Action Type (summarised in table 7.2). Particularly in cases where the lexical level alone does not contain sufficient information to clearly link the variant to one variable, prosody needs to be considered part of the form. While this is not standard in variationist work (see Pichler's (2013b: 45) note on work on prosody), there is a large body of interactional work looking at the prosody of interaction (Couper-Kuhlen and Selting 1996; Barth-Weingarten et al. 2010) that has clearly shown how the same lexical item or vocalisation can do a number of different functions, and that these functions are disambiguated by the prosodic shape of the token (Betz and Golato 2008; Watts 1989; Yaeger-Dror and Hall-Lew 2002; Zellers and Ogden 2013). This illustrates that scholars with a more interactional focus approach interaction as a network of multi-layered cues, and that we need to understand how these cues work in concert rather than creating artificial separations between them.

Interestingly enough, this is similar to part of Schegloff's (1993) critique of CA-based coding and quantification schemes: He argues that we *must not* quantify, because anything in interaction might be potentially relevant, and we therefore cannot make any a-priori decisions about what to include in our description of the variants. Stivers (2015), picking up on this critique, aptly conceptualises coding schemes as knowledge frozen in time. Taken as a cautionary note and applied more broadly to the development of any coding scheme, this perspective can be hugely constructive because it invites us to firstly reflect on and explain both our focus and our categories, and secondly to keep an acute awareness that we might

173

need to revise the coding scheme to include more or different details as we learn more about the phenomenon we are studying.

Based on this previous work, I argue that prosody (here primarily pitch and amplitude) is an integral part of the form for Listener Responses, particularly in multifunctional lexical tokens – as Goffman already noted in his 1955 essay 'On face-work', a 'momentary change in tone of voice' can (quite radically) change the meaning of what is being said. I will show that both the lexical and the prosodic part of the form need to be described relative to the preceding Speaker-talk rather than to the linguistic production of the Listener. This ties the analysis and the theoretical and methodological contributions of this thesis back to the level of actual linguistic realisation, and thus to what is generally treated as language variation. This includes phonetics, lexical choices, morphosyntax, and prosody. Here, the individual action types are considered as the variables, and the *actual linguistic realisation*, i.e. their lexical and prosodic form, are treated as the variants. 'Lexical form' refers to the transcribed utterance, 'prosodic form' to its pitch and amplitude.

This chapter contributes to the description of the variants at a more abstract level by mapping the link between form and function for the individual Action Types both quantitatively and qualitatively. On a quantitative level, this link can be salient based on a simple frequency association, or based on the distinctiveness of one given term or formulation for a specific Action Type. 'Distinctiveness' can be measured by relating how frequent a term is within one Action Type relative to its overall frequency across all Action Types. This exclusivity or typicality of a link is calculated as *term-frequency inverse document frequency (tf-idf)*.[4]

The more common way of looking at the link between form and function is *term frequency* on its own. To give an example, the more often an acknowledgement is done using *yeah*, the more strongly *yeah* is taken to be associated with acknowledgements. The quantitative analysis will explore both these perspectives, with the frequency analysis showing that if we consider lexical form alone, the same 'form' can be used to do different Actions. In the qualitative analysis, I demonstrate that in these cases, prosody is an integral part of the form, being one of the core contextualisation cues that allow interactants to negotiate the 'meaning' (or function) of any given utterance.

---

[4]'Term frequency' rather than the frequency of full utterances was chosen because it allows us to generalise what *types of* words (based on semantic or syntactic characteristics) are associated with each Action Type.

7.1 Analysis

The analysis is structured as follows: I will first present a quantitative analysis showing the most *distinctive* lexical items for each Action Type, using the metric of *term-frequency inverse document frequency*. I will then present the most *frequent* lexical items for each Action Type, followed by the most frequent full formulations. We will see that the same lexical material (words or vocalisations) can be used to do different Actions, even when considering full formulations (ie all the words and vocalisations that constitute the LISTENER RESPONSE in question). In the final subsection of the analysis I will therefore draw on extracts from the conversations presented in chapters 4 and 5 to show that (1) LISTENER RESPONSES are formatted with respect to the preceding talk on all levels of linguistic structure, and (2) especially prosody needs to be described relative to the previous talk. We also see that (3) in multifunctional lexical formulations prosody is an integral part of the form.

7.1.1 *tf-idf*: What each Action Type *typically* looks like

Term frequency inverse document frequency (tf-idf) describes how distinctive a lexical item is to a group of documents, in this case for a specific Action Type. It is a combination of a term's frequency (*tf*) with its *inverse document frequency (idf)*. *Idf* assigns a weight to each word – the more rare the word is in a collection, the higher its weight.

If we multiply *term frequency* with *inverse document frequency* to *tf-idf*, the result is *term frequency* adjusted for how rarely the term is used (Silge and Robinson 2017). This advantage of balancing *term frequency* with the overall frequency of said term across all documents does have the drawback of not showing when the same term is used frequently in several sub-categories. Hence, *tf* and *tf-idf* are best considered in tandem. Figure 7.1 shows *tf-idf* for the different action types. Higher *tf-idf* indicates a stronger unique association between that term and the Action Type. We will look at *term frequency* on its own in the next section of the analysis.

Tf-idf splits the Action Types in to two categories: JOINT CONSTRUCTIONS and QUESTIONS do not seem to be characterised well by specific individual words, while all others are. Nevertheless, we can see patterns in terms of the *type of* lexical material that distinguishes each Action Type from the others.

Starting in the top left facet, ACKNOWLEDGEMENTS are characterised by minimal tokens, some of them vocalisations rather than words, that frequently stand alone (see figure 7.3). JOINT CONSTRUCTIONS are distinguished by prepositions and

FIGURE 7.1
MOST DISTINCTIVE LEXICAL MATERIAL FOR EACH ACTION TYPE: SOME
ACTION TYPES ARE CHARACTERISED WELL BY A SUBSET OF TOKENS, WHILE
OTHERS ARE NOT.

lexical items we would expect to see in syntactically fully-formulated phrases. This suggests the main aspects distinguishing them from all other Action Types are their length and syntactic complexity. Similarly, QUESTIONS (or other-initiations) are characterised by interrogative pronouns, indicating that their syntactic formatting as questions sets them apart.

Tf-idf for WORD SUPPLIES (self-initiated other-repair) looks somewhat puzzling at first: *bugs, mental, up, or,* and *something* do not seem to have much in common. In fact, *bugs* and *mental* are content words (unlike any of the most distinctive tokens

associated with the other Action Types), while the other three are words that can be used to indicate that the utterance is a suggestion. The relatively low *tf-idf* values from JOINT CONSTRUCTIONS, QUESTIONS, and WORD SUPPLIES further remind us that for these actions the specific lexical format of the Listener Response is largely contingent on what the Speaker has said in the preceding TCU (see categorisation of Listener Responses based on how much influence the Speaker and Listener respectively have in chapter 6, as well as the taxonomy of Listener Responses in chapter 5).

Generalisations for the final three Action Types (the right column of facets in figure 7.1) are more intuitive: FIRST ASSESSMENTS are nearly all assessment terms, and SURPRISE MARKS have *oh, okay, ah, wow*, and *right* as the most distinctive tokens, all of which have been described as doing surprise, indicating a 'change-of-state', or serving as a 'news receipt' in past work (see for example Heritage (1984), Heritage (1998), Golato (2012), and Local (1996)). The five most distinctive tokens used to do SECOND ASSESSMENT are terms expressing agreement or intensification.

These observations closely reflect – and expand on – the description of lexical form and overall structure of the individual Listener Response Action Types presented in chapter 5 and summarised in table 5.2.

7.1.2 *term frequency*: What each Action Type *usually* looks like

Now that we know which (types of) lexical material are *typical* for each Action Type, let us take a look at the frequency association – initially based on individual lexical items, and then based on full formulations (ie *all* the lexical material that makes up a given LISTENER RESPONSE) to account for the fact that some Listener Responses are longer than one word.

Lexical items

Figure 7.2 is structured just like figure 7.1, but it shows proportional frequency rather than *tf-idf* on the x-axis.

At the most general level, we can see that ACKNOWLEDGEMENTS, SURPRISE MARKS, and SECOND ASSESSMENTS are fairly well described by the five most frequent lexical items within each group, while this is not the case for the other four Action Types. As mentioned in the last section, this is unsurprising when considering the sequential context and the type of work these different actions do. SELF-INITIATED OTHER-REPAIR (or word-supply) is very brief, and which lexical material is relevant depends entirely on the talk the Speaker has just produced (or is trying to produce). There is no canonical form like the one that has been described for ASSESSMENTS or SURPRISE MARKS (Pomerantz 1984; Ogden 2006; Heritage 1984). QUESTIONS

Most frequent words by action type

FIGURE 7.2

MOST FREQUENT LEXICAL ITEMS FOR EACH ACTION TYPE SHOWING
OVERLAP BETWEEN FORM AND FUNCTION FOR SEVERAL ACTION TYPES.

(other-initiation) and JOINT CONSTRUCTIONS are longer utterances whose lexical and morphosyntactic shape largely based on the preceding TCU. Thus, we would not expect any individual word to be particularly frequent.

With respect to how unique the link between lexical form and Action Type is, this visualisation shows that several tokens which may stand on their own are frequently used to do different actions. These are *yeah, really, right*, and *oh*.[5] I will briefly discuss the individual multifunctional tokens.

---

[5]Note that *you* appears in JOINT CONSTRUCTIONS, QUESTIONS, and WORD SUPPLIES, but it cannot stand alone; for this particular word the collocation it is part of distinguishes these functions.

yeah frequently appears in ACKNOWLEDGEMENTS (42%) and SECOND ASSESSMENTS (19%), often as a lone-standing item, and in JOINT CONSTRUCTIONS and WORD SUPPLIES (4% respectively), but only as part of more elaborate phrases. *Really* appears both in SURPRISE MARKS (10%) and QUESTIONS (5%), while *right* can be an ACKNOWLEDGEMENT (6.5%) or a SURPRISE MARK (10%). *Oh* is the most prototypical SURPRISE MARK (24%), but can also be part of a FIRST ASSESSMENT (4%).[6]

The appearance of the same lexical item in several Action Types does not necessarily have to entail ambiguity in the link between form and function, even on the lexical level alone: as suggested above, these words might be part of a longer utterance, so that the collocation they appear in distinguishes them. The next section shows that these formulations do in fact stand alone and seem to be multifunctional; based on lexical form alone the link between form and function is indeed ambiguous.

Full formulations

Figure 7.3 shows the five most frequent utterances (what I call 'full formulations' here) for each Action Type.[7] Formulations were filtered to only include those that occur more than once within any given action type. JOINT UTTERANCES and WORD SUPPLIES (self-initiated other-repair) have disappeared from this overview, because in these two Action Types no lexical formulation is used twice. This reflects the fact that their lexical and morphosyntactic form is entirely dependent on the preceding talk. A WORD SUPPLY, by definition, needs to supply a word that is a likely continuation of the Speaker's talk, and JOINT CONSTRUCTIONS use lexical material that is part of the ongoing project the Speaker is constructing and extend it. Thus, there is no pre-defined subset of lexical items that could describe these two categories. For the same reason, there are only four formulations given for QUESTIONS; only *really* and *oh really* occurred more than twice, and *is it* and *do you* twice in the questions analysed.

Returning to the broadest level, we can see that two thirds of all ACKNOWLEDGEMENTS (68%) and just about half of all SURPRISE MARKS (46%) are accounted for by the top five formulations, while these add up to less than 15% for each of the other three Action Types presented. Despite this diversity in formulations, the clear association between FIRST ASSESSMENTS and assessment terms, and SECOND

---

[6]All percentages given here refer to the relative frequency of this lexical item *within* this Action Type, *not* the frequency of the Action Type for that lexical item.

[7]'Full formulations' was chosen as a term over 'collocations' or n-grams because those imply a certain number of words. The unit of analysis in this present section is all the vocalised (lexical) material that makes up any given LISTENER RESPONSE.

Top 5 utterances by Action Type

FIGURE 7.3

OVERVIEW OF THE MOST FREQUENT FULL FORMULATIONS FOR EACH
ACTION TYPE, SHOWING THERE IS STILL AMBIGUITY BASED ON LEXICAL
FORM ALONE. JOINT CONSTRUCTIONS AND WORD SUPPLIES ARE NOT
SHOWN, BECAUSE NO FORMULATION IS USED MORE THAN ONCE.

ASSESSMENTS and agreement or intensification noted based on figures 7.1 and 7.2 is evident here, too.

Even when considering the full formulations, there is still ambiguity based on lexical form: *Yeah* occurs as an ACKNOWLEDGEMENT and a SECOND ASSESSMENT, and it can also be a QUESTION (other-initiation), although this is not represented in figure 7.3 because in my data there is only one instance of *yeah* as an OTHER-INITIATION.[8] *Wow* occurs as a MARKER OF SURPRISE and as a FIRST ASSESSMENT, and *(oh) really* occurs in both SURPRISE MARKS (20%) and QUESTIONS (15%) as one of the most frequent formulations for each action type respectively.

Thus, if we treat the lexical material as the 'form' of Listener Responses, there is a relatively high level of ambiguity in the system. This ties in with the decrease in inter-rater reliability for these particular action types noted in the methodology

---

[8]Note that from an interactional point of view it does not matter *how often* something is done – the fact *that* participants do it shows that it is a possible and 'valid' interactional strategy.

(chapter 3). For *yeah* this ambiguity is especially challenging, given that *yeah* accounts for nearly 40% of all ACKNOWLEDGEMENTS and for 10% of all SECOND ASSESSMENTS. It is known as a multifunctional token, also frequently signalling incipient Speakership (see definition of Listener Responses presented here in chapter 4, and the 1993 ROLSI Special Issue (Drummond and Hopper 1993a; Drummond and Hopper 1993b)). In order to understand how participants resolve or avoid this ambiguity in interaction, we need to return to an interactional analysis of the formulations in question.

To summarise, the quantitative analysis presented above has shown that the precise realisation of any given LISTENER RESPONSE in terms of lexical choices, prosody, and morphosyntax, varies based on what this LISTENER RESPONSE is doing. One perspective of describing a form-function link for the individual ACTION TYPES is how *typical* (types of) words or formulations are for that action (*tf-idf*), and another which words are being used particularly *frequently* to do one Action. There are two sets of Action Types: The first group's lexical form varies because these utterances need to be designed to respond exactly to the preceding talk, while the second one is characterised well by a small number of minimal tokens. However, we have also seen that the same form can be used to do different actions. This suggests that 'form' is underspecified if we only include the lexical material, at least for these utterances that can be used to do a number of actions. This brings us back to the introductory quote of this chapter – *how* something is said, especially its prosody, can make all the difference to what it 'means'.

7.1.3   Including Prosody – Qualitative Re-Analysis

In the following section I will demonstrate that there is indeed a clear form-function link if we consider prosody as an integral part of the form for the lexically ambiguous utterances. In order to describe which prosodic cues relate to which function, I present qualitative analyses of representative examples of the ambiguous cases. I will only present one representative example for each of the Action Types where we have seen overlap; most of them have been introduced chapter 5. Examples for ACKNOWLEDGEMENTS, SECOND ASSESSMENTS, OTHER-INITIATIONS, FIRST ASSESSMENTS, and SURPRISE MARKS will be presented in listed order. Table 7.2 summarises these exponents of 'form' for all seven Action Types. Those of the most frequent lexical items per action type that will be discussed in the following have been italicised.

	Lexical shape		Prosodic shape	
	Most *distinctive*	**Most *frequent***	**Pitch**	**Amplitude**
ACKNOWL-EDGEMENT	mhm, uh-huh, okay, mm, laugh	*yeah, wow*, mhm, right, laugh, mm	unmarked	low
SURPRISE MARK	oh, okay, ah, wow, right	*wow*, oh, okay, right, really	rise-fall	matched
FIRST ASSESSMENT	assessment terms	*wow*, yeah, that's, good, oh, it	unmarked	matched
SECOND ASSESSMENT	agreement and intensifying terms	*yeah*, I, it, no, is	unmarked, elongated	matched
WORD-SUPPLY	content words matching the Speaker's needs	yeah, you, to, it, the	sight rise	lower
OTHER-INITIATION (QUESTIONS)	interrogative pronouns	*you*, really, and, it, do	rise	matched
JOINT UTTERANCE	content words, prepositions	yeah, you, to, it, the	matched	matched

TABLE 7.2
LEXICAL AND PROSODIC COMPONENTS OF 'FORM' FOR LISTENER
RESPONSES BY ACTION TYPE

Prosodic design of ACKNOWLEDGEMENTS

The vast majority of ACKNOWLEDGEMENTS are produced in overlap with ongoing talk, as described in chapter 5. Their amplitude is lower than that of the surrounding talk, and they are not elongated or pitch-marked in any way. Example 26 presented and discussed in chapter 5 as extract 10 illustrates this. The *yeah* in line 7 is produced in overlap with the ongoing multi-unit-turn, and the degree signs around it represent its quietness relative to the surrounding talk.

(26)   DONNA AND PUZZLEB, BIRTHDAY PARTIES ON MDI (LINES 2-8 FROM EXTRACT 10, CHAPTER 5)

```
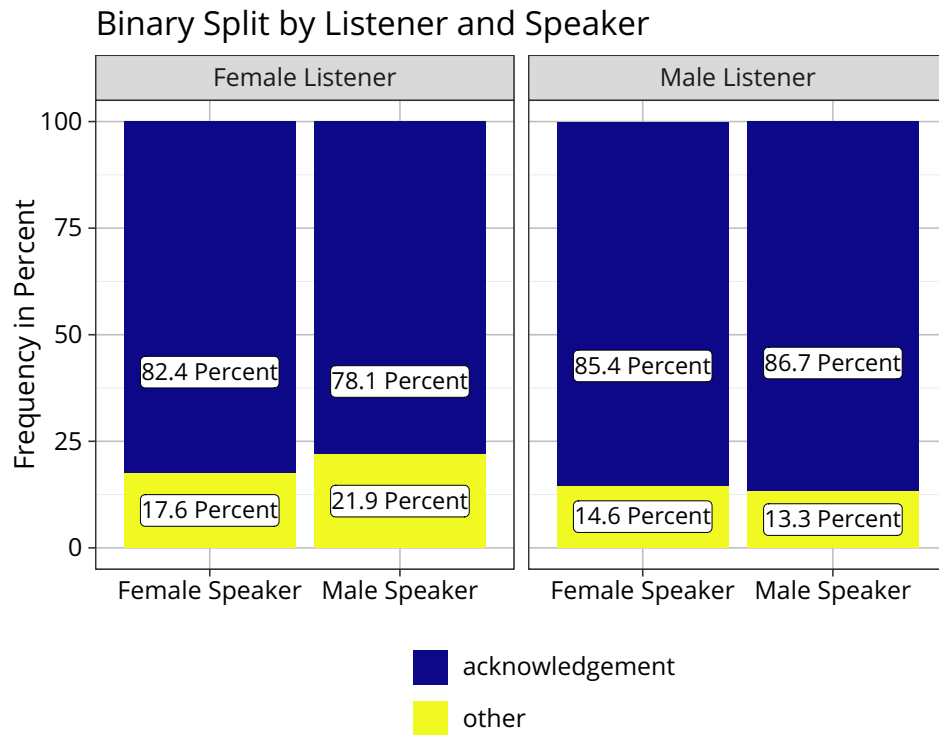2        Donna:            [I used to find that hard when I was young- li-
3                          like-
4        Donna:            children's parties when I was first diagno:sed-
5        Donna:            be[cause it wasn't carb counting.
6        PuzzleB:          [°right°.
```

182

```
7 →    PuzzleB:        [°yeah°,
8      Donna:          [I kinda had to eat the same amount every day,
```

Prosodic design of SECOND ASSESSMENTS

In contrast, SECOND ASSESSMENTS that consist only of *yeah* are prosodically matched to the surrounding talk in the sense that their amplitude is similar. They are elongated to up to 1.4 seconds. This sets them apart from *yeah* as ACKNOWLEDGEMENTS and as OTHER-INITIATIONS.

Most *second assessments* contain some sort of intensifying lexical item that confirms and upgrades the assessment made in the preceding increment. However, some SECOND ASSESSMENTS can be done with only *yeah*. Structurally, the TCU preceding a second assessment contains a first assessment, and very often a tag question or a different strategy for eliciting a confirmatory response. Example 27 presents one of those elongated SECOND ASSESSMENT instances of *yeah* (line 53). The SECOND ASSESSMENT matches the preceding talk in terms of its amplitude and is not produced in overlap with the preceding TCU. It is only PuzzleB's continuation that overlaps with the tail end of the elongated vowel.

(27)   DONNA AND PUZZLEB, DIABETES CAMP AND DAFNE (LINES 51-56 FROM LONG EXTRACT 1, CHAPTERS 4 AND 5)

```
51       PuzzleB:        ↑oh right sounds quite good isn't it that there's still
52                       the:-
53 →     Donna:          yeah:[:-
54       PuzzleB:             [yeah and the-
55       PuzzleB:        yeah cause that's what happened my daughter after going
56                       on the: um camp is that-
```

Though the *yeah* in this particular example is not quite 1.4 seconds long, it is long relative to the speed of the surrounding talk, as we can see in figure 7.4. The *yeah* is 0.5 seconds long, while all of PuzzleB's talk on lines 51 and 52 is uttered within 1.7 seconds, and her response on line 54 in 0.6 seconds. Thus, relative to the surrounding talk, this is quite an elongated realisation of *yeah*.[9] Figure 7.5 shows that there is very little pitch movement on the *yeah*; it starts at 250 Hz, drops to 220 Hz, and then rises to 240 Hz again.

---

[9]The timestamps at the top of the praat pictures here and in the next examples refer to the extract's position in the original recording.

1035.72203   1036.21112

sounds quite good isn't it that there's still the | | Yeah, and the

| | yeah | |

1034

1037

Time (s)

FIGURE 7.4
WAVEFORM AND DURATIONAL INFORMATION OF EXTRACT 27, EXTRACTED
WITH PRAAT.

630

Pitch (Hz)

75

yeah

1036   1036

Time (s)

FIGURE 7.5
PITCH CONTOUR OF THE 'YEAH' ON LINE 53 IN EXAMPLE 27, EXTRACTED
WITH PRAAT.

Prosodic design of OTHER-INITIATIONS

If the OTHER-INITIATION is not formatted as a question with an interrogative particle
(see figure 7.1), it tends to be prosodically marked as one with a pitch-rise, which is
commonly interpreted as interrogative intonation. OTHER-INITIATIONS are prosodi-
cally matched to the surrounding talk in that they are either at a similar or slightly
higher amplitude. This distinguishes particularly *yeah* as OTHER-INITIATION from its
more frequent function as an ACKNOWLEDGEMENT.

184

Lily and Tess, minute 06:54 'Yeah' as a request for elaboration vs 'yeah' as an acknowledgement

```
1       Lily:          I totally get that as a doctor as well-=
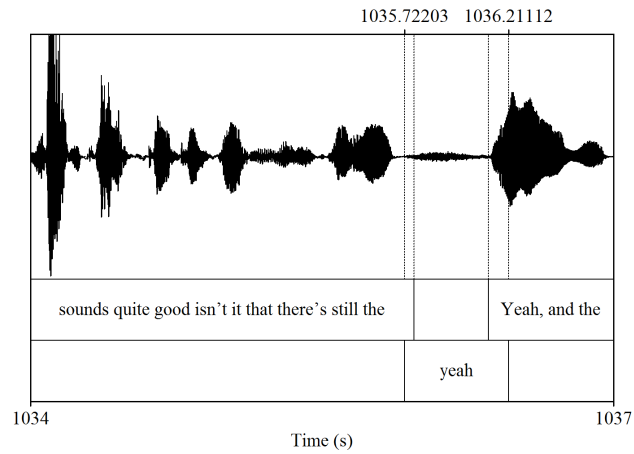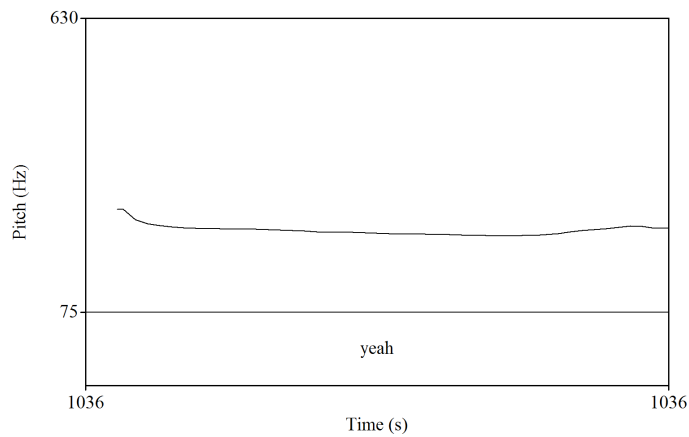
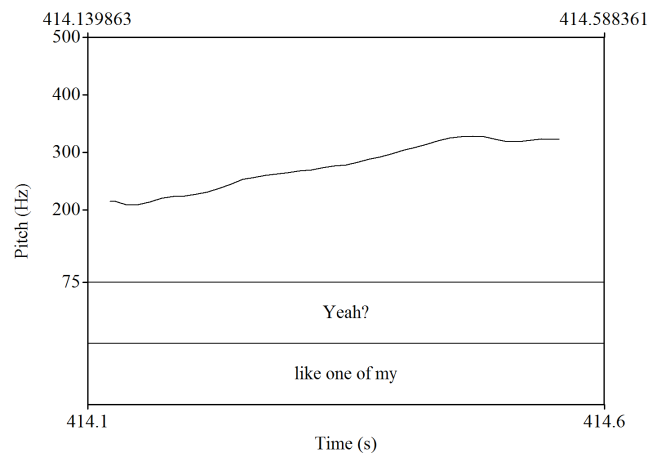2       Lily:          =I get like a total over n- overwhelming paranoi-=

3       Lily:          =it almost became [like one] of my ↑hypo symptoms,

4  →    Tess:                          [↑yeah-  ]

5       Tess:          °yeah,°

6       Lily:          people think I'm an awful person-

7       Lily:          they think I [never pull my weight.

8       Tess:                       [((laughter))

9       Lily:          they think [((inbreath)) like-

10      Tess:                     [↑I know ↑I know.

11      Lily:          um-


               LILY CONTINUES
```

The extract presented above is taken from the conversation between Tess and Lily. They have been discussing symptoms of hypoglycaemia, and just before this extract starts, Tess has described how she feels bad about needing to take a break from work to wait for her blood sugar to rise again after she has done the visible part of 'treating' the hypoglycaemia by eating something. Thus, when Lily describes her own paranoia about other people thinking her lazy when they see her not working but also not obviously 'treating' a hypo (lines 1-3), Tess invites further talk with the OTHER-INITIATION *yeah* (line 4). Lily shows her orientation to this as a repair-initiation (an invitation to elaborate further on her point) by providing more details on the thoughts going through her mind in these situations (lines 6, 7, and 9). The repair-initiation is strongly pitch-marked with the 'question intonation' described earlier, and contrasts with the ACKNOWLEDGEMENT use of *yeah* in line 5.

Praat pictures of the duration and pitch contour of both instances of *yeah* are given below. To create these pitch contours, only the channel with Tess' voice was used, but the tier containing the transcription of Lily's speech in overlap with the first token (here on line 4) was retained. This tier is empty in figure 7.7 because there is no overlap between the response and Lily's ongoing talk.

In the OTHER-INITIATION *yeah* (line 4, figure 7.6), Tess' pitch rises from 215 to 330Hz, in contrast with the shift from 215 to 250Hz in the *yeah* doing ACKNOWL-EDGEMENT on line 5 (figure 7.7). The OTHER-INITIATION is also slightly longer, with a duration of nearly 0.5 seconds compared to 0.4 seconds for the ACKNOWLEDGEMENT that follows it.

FIGURE 7.6
PITCH CONTOUR AND DURATION OF 'YEAH' ON LINE 4, EXTRACTED WITH PRAAT.

FIGURE 7.7
PITCH CONTOUR AND DURATION OF 'YEAH' ON LINE 5, EXTRACTED WITH PRAAT.

Note that, as described above, OTHER-INITIATIONS (questions) can be done with a lone-standing *yeah*, but *really* or *oh really* are far more frequent in my data. The same prosodic pattern applies to all three of them, setting them apart from their use in other sequential contexts. OTHER-INITIATIONS can also be thought of as requests for information or elaboration – they signal some sort of trouble in what the Speaker has said in the preceding TCU.

So far, we have seen *yeah* as a ACKNOWLEDGEMENT, SECOND ASSESSMENT, and as doing OTHER-INITIATION, with prosody being the main cue Listeners and Speakers draw on to distinguish between these functions. In all cases, the prosodic shape in terms of pitch and amplitude needs to be described relative to the pitch and

186

amplitude of the preceding talk produced by the Speaker rather than relative to the Listener's talk in other places in the interaction.

Prosodic design of FIRST ASSESSMENTS

The next two examples will focus on three instances of *wow*: The first extract contains two; one doing FIRST ASSESSMENT, the second ACKNOWLEDGEMENT, and the second extract has one instance of *wow* serving as a SURPRISE MARK.

Very minimal FIRST ASSESSMENTS tend to be prosodically similar to the surrounding talk – they are higher in amplitude than an ACKNOWLEDGEMENT would be, but quieter and less pitch marked than a SURPRISE MARK. The first two uses are exemplified in example 29, the third in example 30 in the next subsection.

(29)   PUZZLEB AND DONNA, MINUTE 06:04, DIABETES THERAPY IN THE 90S

```
 1       Donna:        it was all different types insulin-
 2       Donna:        [there was like a] mixed one-
 3       PuzzleB:      [alright.        ]
 4       Donna:        a long-[acting one and a] fast one,
 5       PuzzleB:             [alright .      ]
 6       PuzzleB:      °wow°.
 7       Donna:        and it was syringes-
 8       Donna:        you had to draw up out the vial and-
 9   →   PuzzleB:      wo[:::w.
10       Donna:          [tap the air bubbles out,

                   DONNA CONTINUES
```

In extract 29 presented above, Donna tells PuzzleB which insulins (lines 1-4) and insulin delivery system (line 7) she had when she was first diagnosed with Type 1 Diabetes in the late 1990s. Different types of insulin are still used today, and the same sub-categorisation applies. PuzzleB's Listener Responses in lines 3, 5, and 6 reflect this – they are ACKNOWLEDGEMENTS; lexically minimal, prosodically unmarked, and stand in overlap with the ongoing talk. The *wow* on line 6 is very quiet, prosodically flat, and uttered quite quickly, clearly marking it as an ACKNOWLEDGEMENT. This is what Donna treats it as, simply continuing her ongoing project of talking about diabetes tools at the time of her diagnosis. Once she gets to the description of *how* she administered the insulin, PuzzleB makes a FIRST ASSESSMENT with the elongated *wo:::w* (line 9). This response is almost 1 second long, and its amplitude is similar to that of the surrounding talk. It does not carry a strong pitch movement, distinguishing it from a SURPRISE MARK as we will see in the

next example. A Listener Response that does FIRST ASSESSMENT does not need to be responded to with a second assessment, the Speaker can also simply continue their ongoing talk. This is what Donna does over the course of the following lines, elaborating further on her diabetes management tools in the first years after diagnosis.

Prosodic design of SURPRISE MARKS

Lexically minimal SURPRISE MARKS tend to carry a stronger rising-falling pitch contour than the same token doing a FIRST ASSESSMENT. The extract given below illustrates this.

(30) TESS AND VELOMINATI, URINE ANALYSIS (LINES 5-8 FROM EXTRACT 11, CHAPTER 5)

```
5        Velominati:    [!] the- there was- there were no glucose monitors-
6        Velominati:    ((that that))-
7 →      Tess:          [/wow\.
8        Velominati:    [so it wasn't an option (.) wasn't available.
```

The full extract is presented and discussed as example 11 in chapter 5. The context of the four lines presented here is Velominati (diagnosed in 1980) talking to Tess (diagnosed in 2004) about Diabetes technology available at the time of his diagnosis. When he mentions that there were not even glucose monitors available at that time, Tess produces a SURPRISE MARK, *wow*. This overlaps with Velominati's continuation. It has a lower amplitude than the surrounding talk and carries a slight prosodic rise-fall, but no strong prosodic marking. It differs from *wow* as a FIRST ASSESSMENT in being more marked – the examples of *wow* as a FIRST ASSESSMENT have a similar amplitude and less extreme pitch-movement.

7.1.4 Summary of the analysis

In summary, the analysis demonstrates that lexical and prosodic realisation are only meaningful in their sequential and interactional context: those LISTENER RESPONSES that are clearly distinguished by their lexical and morphosyntactic format are modelled on the talk in the immediately preceding TCU, and in all cases prosody needs to be considered relative to the prosodic realisation of the preceding TCU. Verbally, we describe it as 'louder'/'quieter'/'similar' or as more/less/similarly pitch-marked, while an acoustic analysis would need to consider the delta between the Speaker's and the Listener's amplitude or pitch.

It further shows that 'form' is a complex combination of cues on different levels of linguistic structure. These levels include which formulations, words, or structures are *distinctive* for each Action Type (figure 7.1), which words or formulations are *frequently* used to do each action (figures 7.2 and 7.3), and in cases where these words can be used to do different Action Types, prosody needs to be considered an integral part of the form.

7.2 Discussion

Overall, this chapter has shown that for Listener Response Action Types as variables, 'form' always needs to be considered in its interactional and sequential context. This means analysing it relative to the talk produced by the Speaker rather than by the Listener. We have further seen that the variants are characterised by a complex interplay rather than a simple addition of lexical and prosodic features.

To introduce this point, Listener Response Action Types were discussed as a discourse-organisational variable in parallel to the well-studied phonological contrast between [p] and [b]. Developing a 'structural' definition of the variable is one of the key contributions of this thesis, and its implications for our quantification of frequency as well as within-category variation have been pointed out in chapters 4 and 6. This final analysis chapter has shown that a structural definition further allows us to look at variation in the precise linguistic (and multimodal, though implementing this analysis shall be left to future studies) realisation of each variant. The levels of 'form' analysed here were the lexical items and vocalisations, as well as pitch and amplitude of each LISTENER RESPONSE. We could also focus on specific phonetic features, or other levels of linguistic structure.

I have suggested in several places that each contribution in an interaction is made to fit the preceding talk and its interactional goal, and that therefore we need to describe the specific form of a LISTENER RESPONSE relative to the other-Speaker talk preceding it. So far, this is a purely qualitative observation supported by

interactional and Conversation Analytic work on different phenomena (Gorisch et al. 2012; Szczepek Reed 2007; Ogden 2006). However, a quantitative analysis would be needed to better understand what exactly the impact of this 'fittedness' to preceding talk of each contribution is to the design of Listener Responses. For this analysis, all responses and the TCU preceding and following them would need to be annotated for pitch (including movement, range, and mean) and amplitude (again, including range and mean) and then compared. This was beyond the scope of the presented PhD thesis, but it would be a very interesting and worthwhile follow up study.

It would be especially interesting to do such an analysis, because this ties the level of for example phonetic variation to the level of the interaction – a current goal of 3rd wave studies of language variation. Thus, analyses of discourse-level and phonological structure could be brought into a fruitful dialogue and further enrich each other. Such analyses would also be very interesting in the context of studies of accommodation (Giles 1973) or priming (Tamminga et al. 2016) and allow us to disentangle how the choice of a specific variant is related to the interaction, the interlocutor, and other factors.

To conclude, this final analysis chapter has shown that a structural definition of the variable allows us to also analyse variation in different levels of 'form', and to better understand the interplay between the many layers that together make up the 'form' of any variant. Such a structural definition of the variable makes it possible to transfer much of the valuable work on phonological structure and phonetic variation to how we approach variation at the level above the phoneme.

Part IV

Discussion

Chapter 8

Discussion

> Ich werde nicht enden zu sagen
> Meine Gedichte sind schlecht.
> Ich werde Gedanken tragen
> Als Knecht.
> Ich werde sie niemals meistern
> Und doch nicht ruhn.
> Soll mich der Wunsch begeistern:
> Es besser zu tun.
>
> *Joachim Ringelnatz* (1910)

[1] So far, our analyses of linguistic variation have not consistently operationalised crucial aspects of the interactions this language occurs in. This PhD thesis has paved the way for a sociolinguistic theory and methodology that integrates interactional and variationist methods. It has demonstrated that we need to orient to interactional structure when defining the variable and the envelope of variation, when describing the variants, and when assessing structural (aka internal) constraints. This interactional orientation also needs to permeate our interpretation of the results.

I will first summarise the analysis as a whole, and then move on to discuss implications for the current approach to defining and delimiting the variable and analysing variation. I will first focus on the general example of LISTENER RESPONSES, and then on questions of language and gender related to this variable. In a next step I extend the model to other discourse-pragmatic variables and finally show how the approach presented can link to work on sociophonetic variation. I

---

[1]Translation of the poem into English (my own): I'll never stop repeating / my poems are bad. / I will carry thoughts / as their servant. / I will never master them / but I will not rest. / I shall be driven by one wish: / to do it better.

acknowledge the limitations of this study and suggest avenues for future research, which have become visible or accessible based on the work presented here.

8.1   Summary of all results

In each of the analysis chapters, we see how an interactional understanding impacts on three different aspects of analysing and interpreting variation, specifically (1) how we define and delimit the variable, (2) how we describe the individual variants, and (3) which structural constraints impact on the variable realisations. I will briefly summarise how these three points apply to each part of the analysis and how variation in Listener Responses patterns with respect to gender, the social variable chosen to demonstrate how this theoretical and methodological contribution can be implemented.

The first analysis chapter, chapter 4, introduces Listener Responses as a variable and treats the *presence* of a vocalised Listener Response as the countable variant.[2] The definition of the variable is rooted in interactional structure: Listener Responses are those vocalisations and utterances that are couched in the other person's ongoing talk. This talk is usually part of a longer interactional project. The primary structural constraint that is used to quantify overall frequency is the amount of talk that is being responded to in any given stretch of Speaking and Listening (usually a multi-unit-turn). At the descriptive level, we see a gender effect, with female participants producing the highest number of Listener Responses when Listening to other women, and men producing the fewest when Listening to another man. Further, we note cross-gender accommodation in mixed dyads. However, once inferential statistics (ZIP) with random intercepts for Listener and dyad are applied, turn length is shown to be the most important predictor for number of Listener Responses in any given turn, while Speaker gender is marginally significant at p = 0.048, and Listener gender is not significant.

Chapter 5 addresses the concern that not all Listener Responses are the same, in the sense that they can do a variety of *actions*. Accordingly, a more complex set of variants is developed: a taxonomy of *Action Types* that can then be used as a coding scheme for further quantification. These *Action Types*, or variants, are defined based on the sequential impact they have, using the next-turn proof procedure. In other words, the variants are based on what the Speaker (i.e. the turn-holder) does in following TCU, and also takes into consideration what the

---

[2]Given that I am focussing on vocalised responses only, attempting to count absences would not only be problematic for the reasons outlined in the section on limitations and future work, but in fact impossible, given that there might be non-vocalised multimodal cues like shifts in gaze and posture that serve as responses and signal continued Listenership.

Listener is responding to (i.e. the action in and format of the preceding TCU). However, a separate study is called for, which should look into the precise actions done in and format of the TCUs preceding LISTENER RESPONSES. This chapter does not include any quantification. Rather, it closely illustrates that and how it is important to consider interactional structure and the Speaker's actions in describing and categorizing the Listener's behaviour.

The next two chapters illuminate how an interactional orientation to language variation helps us understand structural constraints on the variation observed, which have been overlooked so far. Chapter 6, treats LISTENER RESPONSES as the variable, and the individual *Action Types* described in the previous chapter as the variants. The structural constraint on variation that is based in the interaction is how much impact which party interactionally has on any given *Action Type*. In other words, what the Speaker does in the turn *preceding* the LISTENER RESPONSE sometimes strongly constrains what the Listener can do, sometimes not at all, and sometimes the next action is jointly negotiated. This is reflected in inferential statistics: across all categories, Listener gender has a greater impact on variation than Speaker gender, but this varies based on the interactional constraint just outlined. This is interesting because it means that in some cases the gender of the person *receiving* the responses statistically matters more than that of the person *doing* responding. Specifically, speaker gender is an important predictor for those *Action Types* that are strongly influenced by the preceding talk, i.e. talk produced by the Speaker, while Listener gender matters most for those Action Types that are not directly made relevant through the Speaker's preceding TCU. Because those make up more than 80% of all LISTENER RESPONSES, this likely tips the scales towards Listener gender as an important predictor across all subsets of LISTENER RESPONSES.

Chapter 7 investigates the link between the individual LISTENER RESPONSE ACTIONS as the variables, and their *lexical and prosodic realisations* as the variants. Analogous to chapter 6, the preceding TCU is the key structural constraint on the variation observed. Here it is the lexical and prosodic realisation of the preceding talk that impacts on the lexical and prosodic realisation of the LISTENER RESPONSE. Accordingly, the prosodic realisation needs to be described *relative to* the preceding talk, rather than in absolute numbers. Similarly, it would not do justice to the interactional structure to normalise these features relative to each person's own production. This is because the realisations are strongly influenced by the Speaker who is being responded to, as well as the Listener's own baseline.

There is a second learning point from an interactional focus when describing typical forms for each ACTION TYPE: In actual interaction, be it face-to-face or via a telephone, the lexical and the prosodic are inseparable. We cannot hear

words without also noticing pitch, amplitude, and other prosodic features. This contrasts with the common practice in studies of language variation of separating the prosodic from the lexical and other levels of linguistic structure. The analysis illustrates that in lexically ambiguous cases it is crucial to be aware of our 'writing bias' and remember that face-to-face communication is a multimodal, immersive experience. The examples presented illustrate that prosody disambiguates functions when the lexical level alone does not.

Overall, this analysis shows that an interactional approach impacts on all levels of analysis when looking at variation in LISTENER RESPONSES. In particular, it helps uncover and understand which constraints that are rooted in interactional structure impact on the variable realisation. I will now turn to a discussion of the implications of the findings and methodological developments just summarised.

8.2   Implications for work on LISTENER RESPONSES

The theoretical and methodological approach presented in this thesis speaks to several bodies of (socio)linguistic work: First of all, it directly relates to previous work on variation in LISTENER RESPONSES. Secondly, it relates to work on language and gender more broadly. Thirdly, it speaks to Discourse-Pragmatic and other sociolinguistic work theorising the variable above the level of the phoneme and working towards integrating interactional and conversational structure into the analysis, and finally it also has relevance for sociophonetic work aiming for a more interactionally sensitive and accountable analysis.

The chapter-internal discussions have primarily focussed on the contributions to work on LISTENER RESPONSES, and the first part of the discussion will review and consolidate these. The second part focusses on language and gender more broadly, further developing arguments introduced in chapter 6. The discussion ends with the big picture: How do my contributions tie in with the broad challenge of conceptualising variables above the level of the phoneme? I also discuss how such an approach can be useful in sociophonetic analyses of language variation and change.

Defining the Variable

There are two sets of variables which I define and analyse here: the first one in chapters 4 and 6 is LISTENER RESPONSES (variants: *absent/present*, and *seven different Action Types* grouped in different ways respectively). The second one in chapter 7 contains the individual ACTION TYPES (variants: *linguistic realisation*). Bear in mind that the Action Types, which are the variants in chapter 6, become the variables in

chapter 7. These shifts in perspective remind us of the many ways of looking at one and the same phenomenon, and that variables above the level of the phoneme need a 'bespoke analysis' (Waters 2016). These 'bespoke analyses' need to be driven by the specific research questions we set out to answer.

For now, let us consider the impact of defining LISTENER RESPONSES and ACTION TYPES as discourse-organisational variables. Rooting the variable definition in the interactional structure has one key advantage: it bases the coding on the participants' orientations and the interaction itself, rather than interpretations the researcher superimposes on the data. This ties in with current aspirations in sociolinguistics to ensure that coding categories and analyses are not only consistent and reproducible (the high inter coder reliability confirms that the analysis presented fulfils these requirements), but also has emic/internal validity based on what is relevant to the participants in any given interaction or context.

Even though it is not possible to count absences for LISTENER RESPONSES and thereby fulfil Labov's *principle of accountability* in the traditional sense, calculating frequency based on the (amount of) talk that is being responded provides an interactionally relevant way of quantifying that is reproducible and comparable across corpora. This elegantly solves the challenge outlined in Murphy (2012): especially in conversations where one party talks more than the other, quantifying based on the total amount of talk produced in the interaction or corpus will lead to extremely skewed results.

Describing the Variants

Let us now move on to the implications of describing the variants based on inter-actional structure. Describing the individual Action Types as variants of LISTENER RESPONSES (and simultaneously as variables with respect to their precise lexical and multimodal realisation) is a contribution to Conversation Analytic work in the following three ways: Firstly, it pulls together disparate strands of work and shows how they fit together. Secondly, such a structured overview of diverse pieces of work on LISTENER RESPONSES reveals research gaps with respect to specific aspects of the phenomenon. And thirdly, it forms the basis for comparisons of different ways of Listening in different interactional contexts.

Particularly this last point ties in with the contribution to Variationist work of describing the variants based on interactional structure rather than on form. First of all, as stated in the preceding subsection on defining the variable, basing our coding and analysis on the orientations the participants display in the interaction ensures the emic validity of our analysis. This is, again, reflected in the extremely high inter coder reliability – the categories are not superimposed by researcher

intuition but derived from the actions observed. Secondly, previous work has shown that there is no clearly delimited set of forms for Listener Responses, illustrating the necessity of defining the phenomenon based on interactional structure rather than form. For example, it has been shown that smiles (Brunner 1979), nods (Stivers 2008), and other multimodal resources (Barth-Weingarten 2011) can serve as listener responses too. Thirdly and finally, defining these more fine-grained variants of Listener Responses that can then be treated as the variables relative to the actual lexical, morphosyntactic, prosodic realisation, connects the interactional and the traditionally linguistic levels of analysis and allows us to transfer this approach to sociophonetic variation. I will return to this point in the final subsection of the discussion.

My analysis shows that listeners can (a) only do preferred second assessments, that (b) second assessments in the present data are very brief and do not follow the standard format outlined by Pomerantz (1984), and (c) the second coder found it challenging to correctly identify them. Thus, second assessments seem to be a structurally and logically possible but practically rare and opaque category that does not lend itself to larger-scale analyses.

I would like to briefly highlight three key questions for Conversation Analytic work on Listener Responses, which can only be formulated based on such a structured overview of *Action Types* and an analysis of the form-function link for individual Action Types. The first relates to preference organisation: Listener Responses tend to be brief, and by definition do not interrupt the flow of the ongoing talk, which suggests they *must* be preferred responses. Gorisch et al. (2012) have shown that aligning Listener Responses are prosodically matched, while disaligning ones are prosodically different from preceding talk. Does this apply in the same way to preferred vs dispreferred Listener Responses, or does the dispreference marking required necessarily lead to a floor change? Secondly, more work on how Listeners multi-modally mark their responses as (not) aligning or affiliating (see also Stivers (2008)) is needed.

Finally, the issue of speakership incipiency of listener responses has been discussed extensively with respect to *acknowledgements*. Previous work has found that about a quarter of them are *speakership incipient* (Drummond and Hopper 1993a), but we do not have comparable studies for the other action types. Based on Lerner (2004b) the norm for collaborative completions seems to be *speakership incipiency*, and assessments are often used in a *sequence-closing* manner. In order to better understand how speakers negotiate these interactional meanings, future work might look into differences in prosodic design, exact sequential position, lexical choice, or other cues in the interaction, similar to chapter 7.

These interesting questions that interactional scholars could build on can only be formulated once we have a taxonomy or 'coding scheme' for Listener Responses that clearly outlines open questions and challenges to such a coding scheme.

Interactional Structure and the Interlocutor as Internal Constraints

Considering the actions and linguistic production of the interlocutor as internal constraints is a novel and unusual perspective in variationist work. Traditionally, variation is analysed at the level of the community (Labovian macro-social studies), or the individual (stance, indexicality, 3rd wave; see Eckert (2012)).[3] Accordingly, the structural constraints on the variable realisation stem from each participant's own production. Thus, one of the most interesting findings is that social characteristics as well as the behaviour and linguistic production of the interlocutor and the interactional space co-created by both participants are important constraints on variation.

I have proposed to think about Listener Responses relative to the surrounding Speaker-talk rather than talk produced by the Listener. As discussed in chapters 6 and 7, this posits the preceding and following TCU done by the *Speaker* as the structural contexts that impact both on *which* Action gets done, and *how* it is realised. With respect to the relationship between Listener Responses as the variable and *Action Types* as the variants, the most important interactional constraint is the *preceding* context, specifically how much the Speaker's action in the preceding TCU constrains what the Listener can do next. With respect to Action Types as the variable and *specific realisation* as the variants, it is again primarily the preceding Speaker-talk that the Listener shows an orientation to in *how* they format their response.

I will now discuss how these contributions speak to work on variation in Listener Responses related to gender, and how they potentially change our approach to and understanding of this variation. I will then discuss how this approach can be extended to other well-established Discourse-Pragmatic variables.

8.3   Implications for the wider field of sociolinguistics

This thesis also contributes to sociolinguistics more broadly: it exemplifies how an understanding of interaction as a collaborative endeavour changes our inter-

---

[3]The sociolinguistic interview, so common in data collection, is therefore regarded as an exercise in eliciting speech from the interviewee without taking into account the *actions of* and *interaction with* the interviewer.

pretation of variation between groups, based on the social variable of gender. The thoughts and suggestions I voice here are in line with the long tradition of broadly interactional critiques of the 'dominance' and 'difference' approaches to gender, eloquently outlined in Cameron et al. (1989), Holmes (1995), Wilkinson and Kitzinger (2014), and Meyerhoff and Ehrlich (2019).

Integrating interactional and variationist approaches in the way presented in this thesis reframes and develops existing work on LISTENER RESPONSES and gender on five different levels: Firstly, with respect to operationalising overall frequency. Second, in terms of the definition of variants (sub-categories) based on interactional structure rather than form. Third, in terms of quantifying variation in the frequency of these sub-categories. Fourth, it also impacts on how we understand this variation and who we ascribe any given action to. Fifth and finally, it speaks to work that focuses on the lexical or prosodic realisation of LISTENER RESPONSES by gender (or culture or any other social variable, for that matter). The second point, defining the sub-categories or variants (the individual Action Types) based on interactional structure will be picked up in the discussion with respect to established Discourse-Pragmatic variables, the other four will be addressed here.

Operationalising Overall Listener Response Frequency

Two key contributions are made with respect to frequency: First of all, quantifying the number of LISTENER RESPONSES relative to the amount of talk in the turn that is being responded to, and secondly, introducing Zero-inflated Poisson regression models as inferential statistics to abstract away from individual differences.

Quantifying the frequency of LISTENER RESPONSES based on the amount of talk in the turn they are responding to solves the challenge Murphy (2012) so clearly pointed out in her paper: If we do not take into account the distribution of roles and talk, and for example relate the number of LISTENER RESPONSES to the total number of words, the 'frequency' we observe might be extremely skewed by an uneven role distribution. This theoretical and methodological contribution builds on the approach presented by Duncan and Fiske (1977), but to my knowledge has only ever been re-used by Dixon and Foster (1998).

Imagine a conversation between A and B, where both together produce 1000 words; 800 of them are by A, and 200 by B. A does 20 Listener Responses, B 80. If we base our frequency calculation on the total words in the conversation, A has a frequency of 2 Responses per 100 words, while B's would be 8 Responses per 100 words. However, these numbers reflect the *role distribution* rather than how often the respective Listener responds when they are *doing being* a Listener: Relative to the number of words produced by the Speaker, A and B each do 10

Listener Responses per 100 Speaker-words – their frequency is actually the same. Thus, studies like Wong and Kruger (2018), Murphy (2012), and Stubbe (1998) confound how often a Listener *responds* with how much each participant *speaks*. The frequency quantification presented resolves this confound and provides a generalisable and thus comparable approach that can be applied to any conversation under study.

Quantifying Frequency and within-category variation

The second contribution made in this thesis with respect to overall frequency is introducing Zero-inflated Poisson regression models that allow for the inclusion of random effects to quantitative sociolinguistics. This makes it possible to abstract away from the large inter-speaker variability previously noted as a challenge and limitation to the generalisability of their findings (Kogure 2003; Fellegy 1995; Oreström 1983). With Zero-inflated Poisson regression, we have a statistical approach that allows us to abstract away from the individual differences within each group and test how much of the variation observed is likely to actually be due to group differences rather than idiosyncrasies.

The same point applies to analysing variation in the frequency of different *Action Types*. Introducing ZIP models, particularly in a Bayesian framework, allows us as researchers to abstract away from the individual differences between participants or conditions (Dyads) and better understand how much of this variation is due to group differences. While these models are already commonplace in fields like medicine, having even found their way into introductory medical statistics textbooks (Cleophas and Zwinderman 2018b), and increasingly used in some areas of linguistics (Coupé 2018; Rigby et al. 2008; Lo and Andrews 2015), I have not seen them applied in sociolinguistic studies of language variation.

Interpreting Variation in the Action Types

The fourth point, analysing and interpreting variation in these sub-categories is extremely crucial and relates to work on LISTENER RESPONSES as well as other interaction-based variables, and to socially situated analyses of language variation and gender more broadly. As discussed at the end of chapter 6, construing the preceding TCU and thus the other-Speaker-talk as an internal constraint allows us to model the influence of interactional structure and of each participant on the actions done. Only coding for Listener-actions and ascribing them to the Listener alone would belie the co-constructed nature of interaction. Specifically, the statistically measurable impact of Speaker and Listener gender vary based on how much the Listener Response is constrained by what the Speaker does in

the preceding TCU. This reminds us of what is intuitive for interactional scholars: we must not a priori ascribe any action to a single participant but start from the assumption that interaction is collaboratively co-constructed. Let us now review previous work on gender-based variation with this knowledge.

As mentioned in the introductory paragraph to this section, this view builds on the 'dominance' and 'difference' approaches to gender (Cameron et al. 1989; Holmes 1995; Wilkinson and Kitzinger 2014; Meyerhoff and Ehrlich 2019), which remind us that the same 'form' is can be used to do a number of different 'functions'. To understand gender-based variation in these forms we need to take into account what the individual participants are actually *doing*. This applies to Listener Responses too, though here the variable is defined based on equivalence in *interactional structure* rather than form.

Additionally, and going beyond past work on socially situated analyses of language variation and gender, I show in chapters 5 and 6 that from an interactional perspective it is impossible to ascribe actions to any single party a priori. Accordingly, when interpreting the variation we observe, we need to group the different Listener Response *Action Types* based on how much impact the Speaker and the Listener respectively have on their presence. This means re-framing our descriptions from 'This group of Listeners *does* these actions more often' to 'This group of Speakers *provides more opportunities* to do these actions'. Such an approach firmly situates the analysis in the interactional structure and reminds us as analysts that conversations are collaboratively co-constructed and closely coordinated endeavours that all involved parties contribute to.

Form, Function, and work on Gender

This orientation to interaction as a collaborative achievement is also reflected on the level of the lexical and prosodic realisation in two ways. First of all, any contribution in an interaction is designed to be 'fit for purpose' lexically, prosodically, and other other levels of structure (see for example Ogden (2006), Gorisch et al. (2012), and Barth-Weingarten et al. (2010)), and this needs to be considered in our analysis. Future work should look more closely into the relationship between the form of the TCU preceding a Listener Response and the response itself, particularly with respect to lexical choice and prosodic design. It might be the case that there are differences in how male and female Listeners design their responses relative to the Speaker's talk – exploring this would be an exciting avenue for future work.

Secondly, and as pointed out in a vast amount of research beyond Listener Responses, the same formulation can be used to do different actions (Jefferson 1984a; Golato 2012; Barth-Weingarten 2011) and the association between form

and function can vary by gender or culture, or be highly idiosyncratic (Maltz and Borker 1983; Jefferson 1984a). The findings presented in chapter 7 corroborate this observation and invite future research into potential social stratification of form-function relationships for Listener Responses.

8.4   Implications for Discourse-Pragmatic Variation

The methodology proposed here ties in with current developments in the study of Discourse-Pragmatic Variation in several important ways, addressing issues of the variable definition, quantification of open and closed sets of variants, and the big bugbear of multifunctionality. I will discuss these points in turn, illustrating how it can be extended to other variables with the examples of Like and neg-tags. This thesis shows how a stringent integration of interactional and variationist methods tentatively explored in Pichler (2013b) addresses the methodological challenges just mentioned.

Defining variable and variants based on interactional structure

In the discourse-organisational approach I have presented, sequence organisation is treated as the syntagmatic axis – Listener Responses are defined as all the things Listeners can do in a specific interactional slot – and the realisation of the Listener Response (Action) as the paradigmatic axis. We can think of the individual Action Types as the first level of depth of the paradigm, and within each Action Type there is the exact realisation of this action as the second level. This enables us to very clearly delimit the envelope of variation and the variants, and to quantify them in a locally relevant manner (see next subsection).

However, this approach cannot be blindly imposed on a form-based variable like like (D'Arcy 2017) or a variable based on functional or derivational equivalence like quotatives (Buchstaller 2006b; Buchstaller 2014) or Neg-tags (Pichler 2016b). As Waters (2016) points out, Discourse-Pragmatic variables are so diverse that we need 'bespoke' analyses for them. We can thus transfer the interactional orientation to coding for function to these and other variables, and integrate the notion of sequential environment as a conditioning factor. How this applies will be discussed in the section on quantifying different sets of variants.

A locally meaningful frequency normalisation

Another recurring challenge in Discourse-Pragmatic work is quantifying and normalising frequency of occurrence in an accountable and comparable way. For

Listener Responses I have proposed quantifying based on the number of words in the turn that is being responded to; this is the locally and interactionally relevant context. Admittedly, the general critique of a certain level of arbitrariness in word counts outlined by Pichler (2010: 597) still holds, but using the R package `tidytext` and re-usable R scripts makes the way I have arrived at the word counts given here maximally reproducible.

Closer interactional analyses of other discourse-level open set variables like Like and Neg-tags might lead to the development of similarly locally relevant measures. This would increase the emic validity of synchronic and diachronic analyses of variation in these two variables. Given that both variables are usually produced within one Speaker's ongoing discourse, a starting point might be to quantify their frequency based on the length of the turn they occur in. The same argument I have presented with respect to Listener Responses, developing the critique in Murphy (2012), can be applied to variables that are internal to one Speaker's discourse: if we quantify its frequency based on overall words in corpus rather than words produced by the respective Speaker and with respect to that specific variable, we are conflating talking time, interactional roles, and token frequency. Different variables will need individually tailored frequency operationalisations, but an interactional approach will always allow us to better understand and model the locally relevant context for any given variable.

Inferential statistics on open and closed sets of variants

In this thesis, as in many other analyses, quantification proceeds from overall frequency to within-category variation. Overall frequency relies on the presence or absence of the variable, though we can only count the presences. This makes it an 'open set' variable, meaning one where we cannot count all variants and thus have difficulty honouring the principle of accountability formulated by Labov. The second level, within-category variation, is generally considered a 'closed set' of variants, because we can code all of the instances of Listener Responses, Like, or neg-tags for their function or form (see also Pichler (2010)).[4]

Crucially, inferential statistics have (to my knowledge) not been done on the 'open set' variables briefly described above so far. Instead, scholars have compared overall frequencies, with the known issues and limitations (see interim discussion of chapter 4). Introducing interactionally accountable ways of normalising frequency (for example based on same- or other-Speaker talk in the relevant turn)

---

[4]I feel somewhat uneasy presenting this as a closed set, because the absences can still not be included in the analysis, while 'zero-variants' are theoretically one of the possible variants. Hence, we need to be very careful to point out what exactly the variable definition and the set of variants are for each step of the analysis.

allows us to apply Zero-inflated Poisson regressions more broadly. These have the advantage of also including 'zero-variants' and allowing us as researchers to test complex mixed effect structures. Thus, if we were to quantify NEG-TAGS relative to the number of words in the turn they are part of, some turns would not contain any NEG-TAGS at all. This would show up as a 'zero' in the aggregated data, and the ZIP model would then estimate (1) the relationship between the zero and the predictors in the logit model, and (2) the relationship between the observed counts other than zero and the predictors in the count model. There is great potential in extending this approach to statistically modelling open class discourse-pragmatic variables to better understand how they vary both synchronically and diachronically.

Generally speaking, the same point applies to what are usually considered 'closed set' variables, i.e. variation in the distribution of the individual Action Types or functions of the phenomenon whose overall frequency we have quantified in the first step. Here, inferential statistics are already possible and Goldvarb is often the programme of choice. I would like to suggest that ZIP models allow for more flexibility, more complex model structures, and, crucially, for predicting non-occurrences of a variant separately from the number of occurrences.

More importantly, an interaction-based coding paying attention to the preceding and following talk can take into account conditioning factors that govern the variation observed and can explain outcomes that might otherwise seem to correlate with broader macro-social variables. For LISTENER RESPONSES this was the case in chapter 6, where for those (and only those) actions that are more driven by the preceding Speaker-talk Speaker gender is a relevant predictor, while Listener gender has a stronger effect on those responses that Listeners can fairly freely decide (not) to do.

Different variables are embedded in the interaction in different ways and will need the aforementioned bespoke analyses. In interactional coding, the locally relevant preceding and following context needs to be considered. For LIKE and NEG-TAGS in dyadic conversations this could be same-Speaker talk or talk done by the other participant, depending on whether the variable is placed turn-initially, medially, or finally, and depending on who held or then takes the floor.[5] This would need to be the first layer of coding – (1) placement and (2) who produces the preceding and following TCU. Once this has been established, we can code for the *action* done in the respective increments.

Let us take the example of turn-final NEG-TAGS. If there is a systematic co-occurrence pattern of a NEG-TAG following an assessment, and the responding

---

[5]Those questions become increasingly complex, and the interactional orientation increasingly important, as the number of participants in the interaction increases.

increment systematically is the preferred response (Pomerantz 1984), this strongly suggests that NEG-TAGS in this context cue alignment. Over three decades ago, Schiffrin (1987b) already suggested that 'discourse markers' might function as contextualisation cues, and this is a very promising avenue to pursue – some Discourse-Pragmatic variables, maybe only in some positions, might indeed turn out to be signs that help the recipient interpret *how* a given utterance is to be understood (see Gumperz (2015: 315)). Contextualisation cues never come alone – they form complex co-occurrence patterns with other cues, including prosody. If we think of Discourse-Pragmatic variables as potential contextualisation cues, this explains why they *can* but do not *have to* be produced in specific contexts.

Going through all the effort of doing such an interactional coding and analysis for Discourse-Pragmatic variables is worthwhile, because it makes the coding categories emically valid and locally relevant. It also makes it possible to tease apart different layers of coding (see the note on multifunctionality in the next subsection). Furthermore, interactional coding and paying attention to the structure of the interaction can bridge the gap between sociophonetic, discourse-pragmatic, and interactional studies. The final section of the discussion is dedicated to the potential benefits of doing so.

A different perspective on multifunctionality

One of the recurring themes and challenges in work on Discourse-Pragmatic variation (see D'Arcy (2017) and Pichler (2016a)) is the (perceived) multifunctionality of one and the same item. Past approaches have addressed this for example by coding for primary and secondary functions, and Pichler (2010) proposes integrating multifunctionality in the coding scheme and the quantitative analysis.

Based on the coding scheme developed and the analysis presented here, I would like to suggest that this 'multifunctionality' might be due to a lack of differentiation of what in fact are distinct levels of linguistic, interactional, or pragmatic structure. I suggest that it is possible to create coding schemes in which all functions (or variants) on any given level are mutually exclusive. 'Multifunctional' thus becomes 'fulfilling functions on different levels of structure'. For LISTENER RESPONSES, this means that the interactional function (as outlined in chapter 5 and applied in chapters 6 and 7) is one level, while aspects like alignment or affiliation, preference organisation, or epistemic marking are different levels for which we need to code separately (see section on Limitations and Future Research below). Of course we need to then consider patterns of co-occurrence of these variants on different levels, because in actual interaction they often go hand in hand and

work together – but not always in the same way. There is great potential for future research in this line of inquiry.

All these points show how integrating and closely intertwining interactional and quantitative analysis contributes to our understanding of Discourse-Pragmatic variation.

8.5 Implications for Sociophonetic Variation

The majority of this thesis is dedicated to modelling variation *above* the level of the phoneme, specifically of LISTENER RESPONSES, and how to construe variables, variable contexts, and variants, based on interactional structure. The final analysis chapter, chapter 7, returns to the fact that these actions need to be put into words, and these words cannot be heard without also perceiving their pitch, amplitude or rate of speech.

This forms a bridge to studies of sociolinguistic variation with respect to phonetic, lexical, morphosyntactic, or other features. The analysis presented in chapter 7 shows that LISTENER RESPONSES are modelled on the form of the preceding TCU on all levels of linguistic structure, particularly their lexical and prosodic form. Which of the two is more relevant depends on a combination of which action is being done, and the length or morphosyntactic complexity of the response.

These findings tie in with the few studies connecting interactional and variationist thinking, as well as with interactional work much less focussed on language variation and change. Three pioneering studies that I am aware of have shown how interactional structure can be relevant to understanding language variation: Nilsson (2015) shows that Swedish non-dialect speakers converge to their dialect-speaking interlocutors and produce dialect features when format-tying. A similar pattern is reflected in my data in JOINT UTTERANCES or SECOND ASSESSMENTS that can be format-tying to the preceding TCU for example at the lexical or prosodic level. Chakrani (2015), in the same Special Issue of *Language and Communication* on Communication Accommodation Theory, shows how speakers of different Arabic dialects orient to and pick up dialect features on the phonological and lexical levels to joke and poke fun at each other. More recently, Raymond (2018) has demonstrated how speakers of different Spanish varieties show an orientation to the socioindexical value of phonological, lexical, and morphosyntactic features by providing metalinguistic commentary on these features or by adjusting their own speech to accommodate to the features their interlocutor is using.

From a more interactional perspective, Gorisch et al. (2012) show that minimal responses are prosodically matched to the surrounding talk when they align, but mismatched when they do not. This ties in with the broader paradigm of 'prosodic orientation' proposed by Szczepek Reed (2007) (see specifically Szczepek Reed (2009), Szczepek Reed (2010), and Szczepek Reed (2011)) and the aforementioned interactional process of format-tying (Goodwin and Goodwin 1987a).

In concert with these different bodies of work, some of which are only emerging now, my study strongly suggests that studies of language variation and change could benefit from integrating an interaction-based coding scheme in three different ways: First of all, single-case analyses of outliers, unusual variants, and metalinguistic commentary can provide insights into language attitudes and reasons underlying both language change and stability (see especially Nilsson (2015) and Raymond (2018)). Secondly, we can investigate the link between form and function and investigate differences between social groups or individuals if we understand that the same action can be done in different ways and that the same lexical or phonetic form can be used to do different actions (see for example Jefferson (1984a) and the last subsection of the discussion on work related to language variation and gender). Last but not least, we could integrate these interactional functions as mediating factors into our analysis of phonological variation. This would allow us to understand how the sequential and interactional context influences the linguistic realisation we observe, and how individual variants accrue meaning in the interaction. It would also tie in with and complement the existing body of excellent third wave sociolinguistic studies that integrate for example ethnographic fieldwork with sociophonetic analyses (Bucholtz 2001; Eckert 2012; Mendoza-Denton 2008).

8.6   Limitations and Future Research

The contributions just described notwithstanding, we also need to acknowledge three main limitations, which provide excellent starting points for future research. The first two limitations are rooted in the interactional paradigm and are a cautionary note not to generalise to other situations, populations, or 'any' male and female Listener or Speaker. The third acknowledges the complexity of interaction and the impossibility of including every interactional detail or all potential sources of variation. I will discuss these points in turn.

As mentioned in the methodology (chapter 3), the recording situation needs to be acknowledged as a constraint when it comes to the frequency of Listener Responses, the types of Actions done, and the frequency at which they are done.

In CA, 'collections of similar cases' conventionally span different interactional situations in different modalities with varying numbers of participants. It is likely that participants in the interactions I recorded are more cooperative and attentive to their conversation partner's cues than they would be in naturally occurring, day-to-day interactions. Hence, I would like to caution against generalising to other contexts and populations from the patterns observed here.[6]

However, my results align very closely with previous other studies that look at various contexts, including audio and video data, suggesting that the processes observed here are not restricted to dyadic face-to-face interactions in a university setting, set up by a researcher. For example, Bavelas et al. (2000) also observed roughly an 80\20 split between what they termed 'generic' and 'specific' responses, and Duncan and Fiske (1977) note very similar rates of 'backchannels' to those observed here. Both research teams elicited talk and staged interactions in a manner similar to the present study. The Action Types I have presented and coded for are derived from existing CA literature and have been attested in other contexts in qualitative studies. Thus they are clearly not specific to my recording situation. Because the coding scheme presented has not been applied previously, there is no point of comparison in terms of the frequency of Action Types. Future work could look into how the different actions pattern across different contexts and modalities.

The related second concern stems from the variationist paradigm: the sample I work with is not gender-balanced, there are more female than male participants, and in fact only one all-male dyad. I note this throughout the thesis and caution against generalising from the patterns observed here to 'male or female Listeners and Speakers in general'.[7] However, the contribution of this thesis is primarily theoretical and methodological, and gender is simply used to exemplify how a theory and methodology that integrates interactional and variationist analyses can be implemented. Future work could reproduce this analysis on a gender-balanced sample drawn from more diverse recording situations, and apply it to other social variables like socio-economic class or ethnicity.

The third set of limitations relates to the necessity of restricting the scope of coding and analysis to match the time and budgetary constraints of a PhD project. In order to develop the pioneering methodological and theoretical contributions presented here and to show how an interactionally sensitive quantitative analysis of variation in Listener Responses can be implemented, it was necessarily to focus

---

[6]I would like to thank Sue Widdicombe and my supervisors, as well as the EMCA Doctoral Network meeting audience in Edinburgh, May 2018, for discussions about this particular challenge.

[7]Even if the sample was fully gender-balanced, it would not be legitimate to extend the findings to other situations, given the implications of the data collection just described.

the coding and analyses to the vocal modality and the local level of interactional structure.

CA tends to be very critical of quantification in general, with some scholars going as far as claiming it is impossible (Schegloff 1993). The analysis presented here exemplifies the challenges of quantifying interactional processes: any quantification, however fine-grained, necessarily abstracts away from the exact detail of what is happening in the individual interaction. In order to develop a coding scheme, we need to abstract away from the single case and focus on those aspects that are relevant to the question we have set out to answer. I have shown here that despite the need to abstract away from the level of detail available to the interactants, we can develop coding schemes and quantify accountably.[8] To do so, we need to derive the coding categories from a close interactional analysis, and keep an open mind and go back to the data when the coding scheme does not seem to fit. However, it is important to remain aware that any coding scheme reflects theoretical and methodological decisions, and that we will never be able to capture *everything*.

Specifically, it was not possible to include the exact placement of LISTENER RESPONSES in the overall frequency analysis (chapter 4). Doing so would allow for analyses of variation in the 'density' of Listener Responses and for a more precise answer to the question 'How many words (can) pass before Listeners (have to) respond?'. With respect to both overall frequency and the frequency of individual Action Types, three other aspects related to interactional structure would have been interesting to consider: First of all, it would be extremely interesting to code not only the Listener Responses but also all Speaker-turns for the actions they do, to understand how these action-sequences are related on a frequency level. Relatedly, annotating which *acknowledgements* are directly made relevant based on contextualisation cues (for example pitch marking or hesitations) would further refine the analysis of Speaker vs. Listener gender as an interactional constraint on Listener Responses. This would also allow for an analysis similar to the experiment presented by Tolins and Fox Tree (2014), who find that, following a surprise mark, Speakers tend to elaborate on the information just given, while following an acknowledgement Speakers continue with providing new information.

Secondly, considering other levels relevant to the interaction, like alignment, affiliation (Stivers 2008), emotional involvement, or affective stance might be related to both how many and which Listener Response Actions get done, and how they are formatted. For example, Gorisch et al. (2012) have found that aligning

---

[8]I argue this is the case not only based on the theoretical approach and methodological tools used, but also because the statistical results closely reflect what we have been able to describe qualitatively for a long time, but have so far been unable to operationalise for quantitative analysis.

responses tend to be prosodically matched to the surrounding talk, while dis-aligning responses are prosodically different. However, these types of coding become increasingly more challenging to root in the interaction. Instead, they read more 'into' rather than 'out of' the observed behaviours.

Thirdly, which Listener and Speaker actions are relevant is also constrained by larger interactional projects like giving an account, telling a story, or producing a list. Among those, narratives are particularly interesting because they require Listeners to signal that their 'mind is with the Speaker' in a number of complex ways (see Sacks' first Winter 1970 lecture as written up in Sacks (1995)). They have been well-researched from a number of perspectives (a particularly good starting point might be Labov and Waletzky (1966) and Labov (2013)), and there is some work on how Listener Responses vary in the different parts of the narrative (Guardiola et al. 2012) paving the way for further work.

Furthermore, there are two exiting avenues of further research with respect to the relationship between form and function. We have seen in the discussion at the end of chapter 7 that three factors together help cue which action a response is doing: frequency and prototypicality of the lexical material, and the pitch contour. This invites a number of interesting experimental follow-on studies, particularly to try and understand if there is any kind of ranking in these three cues. Experimental studies could, for example, look into the relationship between lexical and prosodic design: how do subjects respond to a LISTENER RESPONSE with the prototypical lexical shape of one action type, combined with the prototypical prosodic shape of another? A study design where participants need to continue a turn or a conversation, similar to Tolins and Fox Tree (2014) would be appropriate and relevant. The second avenue I would like to draw attention to here is the opportunity to further develop our understanding of the multimodal composition of the 'form' of LISTENER RESPONSES – previous work has shown that eye blinks (Hömke et al. 2017), smiles or smiley voice (Barth-Weingarten 2011) and a number of other cues Duncan and Fiske (1977) can be (part of) a LISTENER RESPONSE. Future work could look into how exactly these cues interact with each other and extend the analysis presented beyond the vocal modality.

The discussion of the contributions and implications of my theoretical and methodological propositions, as well as all the avenues of future research just presented illustrate the potential impact and usefulness of the approach I have put forward in this thesis. All the individual steps in the methodology and analysis are ready to be extended to more variables, contexts, and levels of variation.

I hope to have shown in this discussion (1) the contributions made to work on LISTENER RESPONSES specifically, (2) that an interactional orientation to discourse-organisational variables changes how we think about past and present work on lan-

guage and gender, and (3) that the methodology presented can easily be extended to other Discourse-Pragmatic variables, and (4) that it can link to sociophonetic work in providing a coding scheme and a better understanding of how interactional structure can impact on phonetic variation. I have acknowledged the limitations of this thesis imposed by scope and time, and shown a number of exciting avenues of future research for which the present analysis has paved the way.

Chapter 9

Conclusion

> Caminante, no había camino,
> se ha hecho camino al andar.
> –
> *Wanderer, there was no path,*
> *Your walking made the path.*

<div style="text-align: right">

inspired by *Antonio Machado*
(1912)

</div>

To conclude, let me summarise and review the main contributions and impact of this thesis. It contributes to sociolinguistic theory and methodology on three levels: (1) with respect to defining and delimiting the variable and the envelope of variation, (2) in defining the variants, and (3) in acknowledging interactional structure and the behaviour and linguistic production of the interlocutor as important internal constraints on variation. On all levels, the interactional structure and the (linguistic) behaviour of the interlocutor are crucial to our operationalisations. I have discussed these with respect to work on LISTENER RESPONSES, in the broader context of work on Language and Gender, with respect to Discourse-Pragmatic variation more broadly, and then briefly outlined how the approach presented can also impact on sociophonetic studies.

The variable above the level of the phoneme – in this thesis LISTENER RESPONSES – is defined in its interactional sequential context. This means that the envelope of variation and the conditioning factors of variation are to be found in said interactional sequential context, too. For the specific discourse-organisation variable presented here, LISTENER RESPONSES, this means that overall frequency needs to be quantified with respect to characteristics of the talk and the participant that are being *responded to*.

Function-based variants are defined based on the action preceding and following that particular response. Here the next-turn proof-procedure is especially

important. It is a strategy interactionalists use to understand and develop their conversation as it unfolds, and which allows us as analysts to create emically valid coding categories. The next-turn proof-procedure in turn makes it possible to distinguish which party primarily drives individual actions, and this can be included in the statistical analysis.

The introduction of Zero-inflated Poisson regression models, particularly in a Bayesian paradigm, means that we can now (1) run inferential statistics on the frequency of LISTENER RESPONSES with a sensible way of dealing with instances of 'absences', and (2) include complex mixed effect structures in these models. Again, the next-turn proof-procedure and the abovementioned awareness that actions are not produced in isolation help us interpret the patterns we observe.

This awareness of interactional structure and the sequential organisation of interaction allows us to see how the actual linguistic and prosodic realisation of any given action is influenced by the preceding talk. For LISTENER RESPONSES this preceding talk is produced by the *other* person, the turn-holder – just as in all the other steps of the analysis. This can bridge the gap between work on sociophonetic variation and variation at the level above the phoneme, and introduce a locally relevant, interaction-based coding scheme into sociophonetic studies.

We can draw close parallels between the level of phonological structure and the level of discourse-organisational structure. This applies to our conceptualisation of the variable, the envelope of variation, the description of the variants, and the sort of structural constraints that need to be taken into account. And just as phonological structure interacts with morphological, syntactic, prosodic, and other levels of linguistic structure, so does the discourse-organisational structure of sequences with the structure of broader interactional projects. In the case of LISTENER RESPONSES this situates the conditioning factor in the *interaction* and in the *other* person's behaviour and linguistic production.

When extending these findings to work on language and gender, I join other broadly interactional critiques of the 'dominance' and 'difference' approaches to supposedly gender-related language variation (see Cameron et al. (1989), Holmes (1995), Wilkinson and Kitzinger (2014), and Meyerhoff and Ehrlich (2019)). Particularly the analysis in chapter 6 shows that the actions (not) done cannot always be attributed to one person only, nor to the person producing them. Rather, paying attention to the organisation of the interaction, we can better understand how each individual utterance came about and how the two participants co-construct their conversation.

The methodological and theoretical approach presented in this thesis can be extended to well-established Discourse-Pragmatic variables, for example LIKE or NEG-TAGS. It extends Pichler's (2013b) call to use CA tools in the analysis of

Discourse-Pragmatic variation. For LISTENER RESPONSES the talk and actions produced by the Speaker (rather than the Listener) provided the crucial envelope of variation and conditioning factors. Integrating interactional and variationist methods in the analysis of other Discourse-Pragmatic variables, it is important to define their context of occurrence, the conditioning factors, the variants, and the envelope of variation for this specific variable. This means that in some cases same-speaker talk will be the preceding and/or following context, and in other places it might be talk and actions done by the other participant. An interactionally accountable analysis will code for this and benefit from this awareness in the analysis.

The statistical tool of Zero-inflated Poisson regressions (in a Null Hypothesis Testing as well as in a Bayesian paradigm) offers increased flexibility for inferential analyses of variation, and – to my knowledge for the first time – allows researchers to model the frequency of overall occurrence. This has so far only been possible as 'normalised' frequency across a corpus, very often normalised based on 'per million words', which is completely removed from the interactional reality of the variable realisation.

In summary, in this thesis I introduce a way of integrating interaction and quantitative analyses of language variation (and by extension change), based on gender-related variation in LISTENER RESPONSES. I demonstrate that interactional structure matters on all levels of analysis, from the definition of the variable to quantifying its frequency, describing its variants, analysing their relative frequencies, and understanding variation in the actual linguistic realisation of the individual LISTENER RESPONSES. At all levels, the actions and linguistic production of the Speaker, i.e. the person holding the floor – not the Listener, i.e. the person doing the response – are crucial in shaping *how often* and *which* Responses get done, as well as *how* they are done. The theoretical and methodological approach I have developed based on LISTENER RESPONSES can be extended to other discourse-level variables, as well as to sociophonetic variation.

Part V

Back matter

Bibliography

Antaki, C., H. Houtkoop-Steenstra, and M. Rapley (2000). "Brilliant. Next Question.'
High-Grade Assessment Sequences in the Completion of Interactional Units'.
In: *Research on Language & Social Interaction* 33.3, pp. 235–262.

Atkinson, J. M. and J. Heritage (1984). *Structures of Social Action: Studies in Conversation Analysis*. Cambridge: Cambridge University Press.

Baayen, R. H. (2008). *Analyzing linguistic data: A practical introduction to statistics using R*. Vol. 2. 3. Cambridge University Press, p. 353. arXiv: `arXiv:1011.1669v3`.

Barnes, S. (2018). 'Copula variation in Asturian Spanish and the multidimensionality of stancetaking in interaction'. In: *Journal of Sociolinguistics* 22.1, pp. 29–54.

Barth-Weingarten, D. (2011). 'Response tokens in interaction prosody, phonetics and a visual aspect of German JAJA'. In: *Gesprächsforschung - Online-Zeitschrift zur verbalen Interaktion* 12, pp. 301–370.

Barth-Weingarten, D., E. Reber, and M. Selting (2010). *Prosody in Interaction*. Vol. 23. John Benjamins.

Bavelas, J. B., L. Coates, and T. Johnson (2000). 'Listeners as co-narrators.' In: *Journal of Personality and Social Psychology* 79.6, pp. 941–952.

Bavelas, J. B., L. Coates, and T. Johnson (2002). 'Listener responses as a communication process: The role of gaze'. In: *Journal of Communication* 52.3, pp. 566–580.

Betz, E. and A. Golato (2008). 'Remembering Relevant Information and Withholding Relevant Next Actions: The German Token achja'. In: *Research on Language & Social Interaction* 41.1, pp. 58–98.

Bilous, F. R. and R. M. Krauss (1988). 'Dominance and accommodation in the conversational behaviours of same- and mixed-gender dyads'. In: *Language & Communication* 8.34, pp. 183–194.

Boersma, P. and D. Weenink (2016). *Praat: doing phonetics by computer*.

Brooks, M. E., K. Kristensen, K. J. van Benthem, A. Magnusson, C. W. Berg, A. Nielsen, H. J. Skaug, M. Mächler, and B. M. Bolker (2017). 'glmmTMB balances speed and flexibility among packages for zero-inflated generalized linear mixed modeling'. In: *The R Journal* 9.2, pp. 378–400.

Brunner, L. J. (1979). 'Smiles can be back channels'. In: *Journal of Personality and Social Psychology* 37.5. Ed. by A. G. Greenwald, pp. 728–734.

Bucholtz, M. (2001). 'The Whiteness of Nerds: Superstandard English and Racial Markedness'. In: *Journal of Linguistic Anthropology* 11.1, pp. 84–100.

Buchstaller, I. (2006a). 'Diagnostics of age-graded linguistic behaviour: The case of the quotative system'. In: *Journal of Sociolinguistics* 10.1, pp. 3–30.

— (2006b). 'Social stereotypes, personality traits and regional perception displaced: Attitudes towards the 'new' quotatives in the U.K.' In: *Journal of Sociolinguistics* 10.3, pp. 362–381.

— (2008). 'The localization of global linguistic variants'. In: *English World-Wide* 29.1, pp. 15–44.

— (2009). 'The quantitative analysis of morphosyntactic variation: Constructing and quantifying the denominator'. In: *Linguistics and Language Compass* 3.4, pp. 1010–1033.

— (2011). 'Quotations across the generations: A multivariate analysis of speech and thought introducers across 5 decades of Tyneside speech'. In: *Corpus Linguistics and Linguistic Theory* 7.1, pp. 59–92.

— (2014). *Quotatives: New Trends and Sociolinguistic Implications*. Oxford, UK: John Wiley & Sons, Inc, p. 306.

Buckland, M. and F. Gey (1994). 'The relationship between Recall and Precision'. In: *Journal of the American Society for Information Science* 45.1, pp. 12–19. arXiv: `arXiv:1011.1669v3`.

Bürkner, P.-C. (2017). 'brms: An R Package for Bayesian Multilevel Models Using Stan'. In: *Journal of Statistical Software* 80.1, pp. 1–28.

— (2018). 'Advanced Bayesian Multilevel Modeling with the R Package brms'. In: *The R Journal* 10.1, pp. 395–411.

Cameron, D., F. McAlinden, and K. O'Leary (1989). 'Lakoff in context: the form and function of tag questions'. In: *Women in their speech communities : new perspectives on language and sex*. Ed. by J. Coates and D. Cameron. London: Longman, pp. 74–93.

Carpenter, B., A. Gelman, M. D. Hoffman, D. Lee, B. Goodrich, M. Betancourt, M. Brubaker, J. Guo, P. Li, and A. Riddell (2017). 'Stan : A Probabilistic Programming Language'. In: *Journal of Statistical Software* 76.1, pp. 1–32.

Cathcart, N., J. Carletta, and E. Klein (2003). 'A shallow model of backchannel continuers in spoken dialogue'. In: *Proceeding EACL '03 Proceedings of the tenth conference on European chapter of the Association for Computational Linguistics* 1, pp. 51–58. arXiv: `arXiv:1011.1669v3`.

Chakrani, B. (2015). 'Arabic interdialectal encounters: Investigating the influence of attitudes on language accommodation'. In: *Language & Communication* 41, pp. 17–27.

Clancy, P. M., S. A. Thompson, R. Suzuki, and H. Tao (1996). 'The conversational use of reactive tokens in English, Japanese, and Mandarin'. In: *Journal of Pragmatics* 26.3, pp. 355–387.

Cleophas, T. J. and A. H. Zwinderman (2018a). 'Bayesian Poisson Rate Analysis'. In: *Modern Bayesian Statistics in Clinical Research*. Cham: Springer International Publishing, pp. 101–110.

— (2018b). *Modern Bayesian Statistics in Clinical Research*. Cham: Springer International Publishing.

Clift, R. (2016). 'Don't make me laugh: Responsive laughter in (dis)affiliation'. In: *Journal of Pragmatics* 100, pp. 73–88.

Cohen, J. (1960). 'A Coefficient of Agreement for Nominal Scales'. In: *Educational and Psychological Measurement* 20.1, pp. 37–46.

Cohen, S., T. Kamarck, and R. Mermelstein (1983). 'A Global Measure of Perceived Stress'. In: *Journal of Health and Social Behavior* 24.4, pp. 385–396.

Coupé, C. (2018). 'Modeling Linguistic Variables With Regression Models: Addressing Non-Gaussian Distributions, Non-independent Observations, and Nonlinear Predictors With Random Effects and Generalized Additive Models for Location, Scale, and Shape'. In: *Frontiers in Psychology* 9, p. 513.

Couper-Kuhlen, E. and M. Selting (1996). *Prosody in Conversation: Interactional Studies*. Cambridge University Press.

D'Arcy, A. (2017). *Discourse-pragmatic variation in context: Eight hundred years of like*. Amsterdam: John Benjamins.

Dines, E. R. (1980). 'Variation in Discourse: And Stuff like That'. In: *Language in Society* 9.1, pp. 13–31.

Dixon, J. A. and D. H. Foster (1998). 'Gender, social context, and backchannel responses'. In: *Journal of Social Psychology* 138.1, pp. 134–136.

Drager, K. and J. Hay (2012). 'Exploiting random intercepts: Two case studies in sociophonetics'. In: *Language Variation and Change* 24.1, pp. 59–78.

Drummond, K. and R. Hopper (1993a). 'Back Channels Revisited: Acknowledgment Tokens and Speakership Incipiency'. In: *Research on Language and Social Interaction* 26.2, pp. 157–177.

— (1993b). 'Some Uses of Yeah'. In: *Research on Language and Social Interaction* 26.2, pp. 203–212.

Dubois, B. L. and I. Crouch (1975). 'The question of tag questions in women's speech: they don't really use more of them, do they?'. In: *Language in Society* 4.3, p. 289.

Duncan, S. and D. W. Fiske (1977). *Face-to-face interaction: Research, methods, and theory*. Vol. 3. L. Erlbaum Associates, p. 361.

Duncan, S. and G. Niederehe (1974). 'On signalling that it's your turn to speak'. In: *Journal of Experimental Social Psychology* 10.3, pp. 234–247.

Eckert, P. (1989). *Jocks and Burnouts: Social Categories and Identity in High School.* New York: Teachers College Press, p. 195.

— (2009). 'Affect, Sound Symbolism, and Variation'. In: *University of Pennsylvania Working Papers in Linguistics* 15.2, pp. 70–80.

— (2011). 'Language and Power in the Preadolescent Heterosexual Market'. In: *American Speech* 86.1, pp. 85–97.

— (2012). 'Three Waves of Variation Study: The Emergence of Meaning in the Study of Sociolinguistic Variation'. In: *Annual Review of Anthropology* 41.1, pp. 87–100.

Eckert, P. and S. McConnell-Ginet (2003a). 'Constructing, deconstructing and reconstructing gender'. In: *Language and Gender*. Ed. by P. Eckert and S. McConnell-Ginet. Cambridge University Press, pp. 9–51.

— (2003b). *Language and Gender*. Cambridge: Cambridge University Press, p. 366.

— (2003c). 'Linking the linguistic to the social'. In: *Language and Gender*. Ed. by P. Eckert and S. McConnell-Ginet. Cambridge University Press, pp. 52–90.

Edelsky, C. and K. Adams (1990). 'Creating Inequality: Breaking the Rules in Debates'. In: *Journal of Language and Social Psychology* 9.3, pp. 171–190.

Eiswirth, M. (2014). 'The Prosody of Epistemic Authority and Subordination in Agreements'. MSc. University of Edinburgh, p. 44.

Fellegy, A. M. (1995). 'Patterns and Functions of Minimal Response'. In: *American Speech* 70.2, pp. 186–199.

Flege, J. E. and W. S. J. Brown (1982). 'The voicing contrast between English /p/ and /b/ as a function of stress and position-in-utterance'. In: *Journal of Phonetics* 10, pp. 335–345.

Ford, C. E. and S. A. Thompson (1996). 'Interactional units in conversation: syntactic, intonational, and pragmatic resources for the management of turns'. In: *Interaction and grammar*. Ed. by E. Ochs, E. A. Schegloff, and S. A. Thompson. Cambridge: Cambridge University Press, pp. 134–184.

Gagnon, D. R., S. Doron-LaMarca, M. Bell, T. J. O'Farrell, and C. T. Taft (2008). 'Poisson regression for modeling count and frequency outcomes in trauma research'. In: *Journal of Traumatic Stress*. Vol. 21. 5. John Wiley & Sons, Ltd, pp. 448–454.

Gardner, R. (1997). 'The conversation object Mm: A weak and variable acknowledging token'. In: *Research on Language and Social Interaction* 30.2, pp. 131–156.

Gardner, R. (1998). 'Between Speaking and Listening: The Vocalisation of Under-standings'. In: *Applied Linguistics* 19.2, pp. 204–224.

— (2001). *When listeners talk : response tokens and listener stance*. J. Benjamins Pub, p. 289.

Giles, H. (1973). 'Accent Mobility: A Model and some Data'. In: *Anthropological Linguistics* 15.2, pp. 87–105.

Goffman, E. (1972). *Encounters: two studies in the sociology of interaction*. Penguin, p. 134.

— (1978). 'Response Cries'. In: *Language* 54.4, p. 787.

— (1981). *Forms of talk*. Oxford: Blackwell, p. 335.

— (1982). *Interaction ritual: essays on face-to-face behavior*. Doubleday, p. 270.

Golato, A. (2012). 'German oh: Marking an Emotional Change of State'. In: *Research on Language and Social Interaction* 45.3, pp. 245–268.

Golato, A. and Z. Fagyal (2008). 'Comparing Single and Double Sayings of the German Response Token ja and the Role of Prosody: A Conversation Analytic Perspective'. In: *Research on Language & Social Interaction* 41.3, pp. 241–270.

Goodwin, C. (1984). 'Notes on story structure and the organization of participation'. In: *Structures of Social Action*. Ed. by J. M. Atkinson. Cambridge: Cambridge University Press. Chap. 10, pp. 225–246.

— (1986a). 'Between and within: Alternative Sequential Treatments of Continuers and Assessments'. In: *Human Studies* 9.2/3, pp. 205–217.

— (1986b). 'Between and Within: Alternative Treatments of Continuers and Assessment'. In: *Human Studies* 9.2, pp. 205–217.

Goodwin, M. H. and C. Goodwin (1987a). 'Children's arguing'. In: *Language, gender, and sex in comparative perspective*. Ed. by S. U. Philips, S. Steele, and C. Tanz. Cambridge: Cambridge University Press. Chap. 8, pp. 200–248.

— (1987b). 'Concurrent Operations on Talk'. In: *Pragmatics* 1.1, pp. 1–54.

Gorisch, J., B Wells, and G. J. Brown (2012). 'Pitch Contour Matching and Interactional Alignment across Turns: An Acoustic Investigation'. In: *Language and Speech* 55.1, pp. 57–76.

Guardiola, M. and R. Bertrand (2013). 'Interactional convergence in conversational storytelling: When reported speech is a cue of alignment and/or affiliation'. In: *Frontiers in Psychology* 4.OCT, pp. 1–17.

Guardiola, M., R. Bertrand, R. Espesser, and S. Rauzy (2012). 'Listener's responses during storytelling in French conversation'. In: *INTERSPEECH*.

Gumperz, J. J. (2015). 'Interactional Sociolinguistics A Personal Perspective'. In: *The Handbook of Discourse Analysis*. Hoboken, NJ, USA: John Wiley & Sons, Inc., pp. 309–323.

Guthrie, A. M. (1997). 'On the systematic deployment of Okay and Mmhmm'. In: *Pragmatics* 7.3, pp. 397–415.

Guzman, E., R. Alkadhi, and N. Seyff (2016). 'A Needle in a Haystack: What Do Twitter Users Say about Software?' In: *Proceedings - 2016 IEEE 24th International Requirements Engineering Conference, RE 2016*. IEEE, pp. 96–105.

Hall-Lew, L., R. Friskney, and J. M. Scobbie (2017). 'Accommodation or political identity: Scottish members of the UK Parliament'. In: *Language Variation and Change*.

Hayes, A. F. and K. Krippendorff (2007). 'Answering the Call for a Standard Reliability Measure for Coding Data'. In: *Communication Methods and Measures* 1.1, pp. 77–89.

Heinz, B. (2003). 'Backchannel responses as strategic responses in bilingual speakers' conversations'. In: *Journal of Pragmatics* 35.7, pp. 1113–1142.

Heldner, M., J. Edlund, and J. Hirschberg (2010). 'Pitch similarity in the vicinity of backchannels'. In: *INTERSPEECH*, pp. 3054–3057.

Hellwig, B., J. Geerts, A. Somasundaram, M. Tacchetti, M. Hulsbosch, and D. van Uytvanck (2018). *Manual for ELAN-Linguistic Annotator: version 5.4*. Tech. rep. Nijmegen: The Language Archive, MPI for Psycholinguistics.

Heritage, J. (1984). 'A change-of-state token and aspects of its sequential placement'. In: *Structures of social action: Studies in conversation analysis*. Ed. by J. M. Atkinson and J. Heritage. Cambridge: Cambridge University Press. Chap. 13, pp. 299–345.

— (1998). 'Oh prefaced responses to inquiry'. In: *Language in Society* 27.3, pp. 291–334.

— (1999). 'Conversation Analysis at Century's End: Practices of Talk-in-Interaction, Their Distributions, and Their Outcomes'. In: *Research on Language & Social Interaction* 32.1, pp. 69–76.

Hills, P. and M. Argyle (2002). 'The Oxford Happiness Questionnaire: a compact scale for the measurement of psychological well-being'. In: *Personality and Individual Differences* 33.7, pp. 1073–1082.

Hoey, E. M. and K. H. Kendrick (2017). 'Conversation Analysis'. In: *Research Methods in Psycholinguistics and the Neurobiology of Language: A Practical Guide*. Ed. by A. M. B. de Groot and P. Hagoort. Wiley-Blackwell. Chap. 8, pp. 151–173.

Holmes, J. (1984). 'Hedging your bets and sitting on the fence: Some evidence for hedges as support structures'. In: *Te Reo* 27, pp. 47–62.

— (1995). *Women, men and politeness*. Longman, p. 254.

Holmes-Elliott, S. and E. Levon (2017). 'The substance of style : Gender, social class and interactional stance in /s/-fronting in southeast England'. In: *Linguistics* 55.5, pp. 1045–1072.

Hömke, P., J. Holler, and S. Levinson (2017). 'Eye Blinking as Addressee Feedback in Face-To-Face Conversation'. In: *Research on Language and Social Interaction* 50.1, pp. 54–70.

Hripcsak, G. and A. S. Rothschild (2005). 'Agreement, the F-measure, and reliability in information retrieval'. In: *Journal of the American Medical Informatics Association* 12.3, pp. 296–298.

Ito, R. and S. Tagliamonte (2003). 'Well weird, right dodgy, very strange, really cool: Layering and recycling in English intensifiers'. In: *Language in Society* 32.2, pp. 257–279.

Jackson, C., J. H. Mostowy, H. R. Stagg, I. Abubakar, N. Andrews, and T. A. Yates (2016). 'Working conditions and tuberculosis mortality in England and Wales, 1890-1912: A retrospective analysis of routinely collected data'. In: *BMC Infectious Diseases* 16.1, p. 215.

Jefferson, G. (1978). 'Studies in the organization of conversational interaction'. In: *Studies in the organization of conversational interaction*. Ed. by J. Schenkein. New York: Academic Press, pp. 213–48.

— (1984a). 'Notes on a systematic deployment of the acknowledgement tokens Yeah; and Mm Hm'. In: *Paper in Linguistics* 17.2, pp. 197–216.

— (1984b). 'Transcript notation'. In: *Structures of Social Action*. Ed. by J. M. Atkinson and J. Heritage. Cambridge: Cambridge University Press, pp. ix–xvi.

— (1993). 'Caveat Speaker: Preliminary Notes on Recipient Topic-Shift Implicature'. In: *Research on Language and Social Interaction* 26.1, pp. 1–30.

Kendrick, K. H. (2017). 'Using Conversation Analysis in the Lab'. In: *Research on Language and Social Interaction* 50.1, pp. 1–11.

Kjellmer, G. (2009). 'Where do we backchannel?: On the use of \textit{mm, mhm, uh-huh} and such like'. In: *International Journal of Corpus Linguistics* 14.1, pp. 81–112.

Kogure, M. (2003). 'Gender differences in the use of backchannels: Do Japanese men and women accommodate to each other?' PhD thesis. ProQuest Dissertations Publishing.

Krippendorff, K. (2004a). *Content analysis: an introduction to its methodology*. 2nd. Thousand Oaks: Sage Publications, p. 413.

Krippendorff, K. (2004b). 'Reliability in Content Analysis.' In: *Human Communication Research* 30.3, pp. 411–433.

— (2011). *Computing Krippendorff 's Alpha-Reliability*. Tech. rep. University of Pennsylvania, p. 12.

Labov, W. (1963). 'The Social Motivation of a Sound Change'. In: *WORD* 19.3, pp. 273–309.

Labov, W. (1966a). 'The linguistic variable as a structural unit'. In: *Washington Linguistics Review* 3, pp. 4–22.

— (1966b). *The social stratification of English in New York City*. Washington D.C.: Center for Applied Linguistics, p. 655.

— (1972). *Sociolinguistic patterns*. University of Pennsylvania Press, p. 344.

— (2013). *The language of life and death: The transformation of experience in oral narrative*. Cambridge: Cambridge University Press, pp. 1–239.

Labov, W. and J. Waletzky (1966). 'Narrative Analysis: Oral Versions of Personal Experience'. In: *Proceedings of the 1966 Annual Spring Meeting of the American Ethnological Society*. Ed. by J. Helm. Seattle and London: University of Washington Press, pp. 12–43.

Lakoff, R. T. (1973). 'Language and woman's place'. In: *Language in Society* 2, pp. 45–80.

Lavandera, B. R. (1978). 'Where does the the sociolinguistic variable stop'. In: *Language in Society* 7, pp. 171–182.

Lerner, G. H. (1995). 'Turn Design and the Organization of Participation in Instructional Activities'. In: *Discourse Processes* 19.1, pp. 111–131.

— (1996). 'On the semi-permeable character of grammatical units in conversation: conditional entry into the turn space of another speaker'. In: *Interaction and grammar*. Ed. by E. Ochs, E. A. Schegloff, and S. A. Thompson. Cambridge: Cambridge University Press, pp. 238–276.

— (2004a). 'Collaborative turn sequences'. In: *Conversation analysis: Studies from the first generation*, pp. 225–256.

— (2004b). 'On the place of linguistic resources in the organization of talk-in-interaction: Grammar as action in prompting a speaker to elaborate'. In: *Research on Language and Social Interaction* 37.2, pp. 151–184.

Levin, H. and T. Lin (1988). 'An accommodating witness'. In: *Language & Communication* 8.34, pp. 195–197.

Levitan, R., S. Benus, A. Gravano, and J. Hirschberg (2015). 'Entrainment and turn-taking in human-human dialogue'. In: *AAAI Spring Symposium on Turn-Taking and Coordination in Human-Machine Interaction*.

Lindström, A. and T. Heinemann (2009). 'Good Enough: Low-Grade Assessments in Caregiving Situations'. In: *Research on Language & Social Interaction* 42.4, pp. 309–328.

Llamas, C., D. Watt, and D. E. Johnson (2009). 'Linguistic Accommodation and the Salience of National Identity Markers in a Border Town'. In: *Journal of Language and Social Psychology* 28.4, pp. 381–407.

Lo, S. and S. Andrews (2015). 'To transform or not to transform: using generalized linear mixed models to analyse reaction time data'. In: *Frontiers in Psychology* 6, p. 1171.

Local, J. (1996). 'Conversational phonetics: some aspects of news receipts in everyday talk'. In: *Prosody in Conversation*. Ed. by E. Couper-Kuhlen and M. Selting. Cambridge: Cambridge University Press, pp. 177–230.

— (2003). 'Variable domains and variable relevance: interpreting phonetic exponents'. In: *Journal of Phonetics* 31.3-4, pp. 321–339.

Local, J. and J. Kelly (1986). 'Projection and Silences: notes on a phonetic and conversational structure'. In: *Human Studies* 9.2-3, pp. 185–204.

Local, J. and G. Walker (2008). 'Stance and affect in conversation: On the interplay of sequential and phonetic resources'. In: *Text & Talk - An Interdisciplinary Journal of Language, Discourse Communication Studies* 28.6, pp. 723–747.

Local, J., J. Kelly, and W. H. G. Wells (1986). 'Towards a phonology of conversation: turn-taking in Tyneside English'. In: *Journal of Linguistics* 22.02, p. 411.

Macken, M. A. and D. Barton (1980). 'The acquisition of the voicing contrast in English: a study of voice onset time in word-initial stop consonants'. In: *Journal of Child Language* 7.01, pp. 41–74.

Maltz, D. N. and R. A. Borker (1983). 'A cultural approach to malefemale miscommunication'. In: *Language and social identity*. Ed. by J. J. Gumperz. Cambridge: Cambridge University Press, pp. 196–216.

Maschler, Y. and D. Schiffrin (2015). 'Discourse Markers Language, Meaning, and Context'. In: *The Handbook of Discourse Analysis*. Hoboken, NJ, USA: John Wiley & Sons, Inc., pp. 189–221.

Maynard, D. W. (2012). 'Everyone and No One to Turn to: Intellectual Roots and Contexts for Conversation Analysis'. In: *The Handbook of Conversation Analysis*. Chichester, UK: John Wiley & Sons, Ltd, pp. 9–31.

Maynard, S. K. (1990). 'Conversation management in contrast: Listener response in Japanese and American English'. In: *Journal of Pragmatics* 14.3, pp. 397–412.

McCarthy, M. (2003). 'Talking back: "Small" interactional response tokens in everyday conversation'. In: *Research on Language and Social Interaction* 36.1, pp. 33–63.

Mendoza-Denton, N. (2008). *Homegirls: Language and Cultural Practice Among Latina Youth Gangs*. Ed. by N. Mendoza-Denton. Oxford, UK: Blackwell Publishing Ltd, pp. 1–339. arXiv: `arXiv:1011.1669v3`.

Meyerhoff, M. (2014). 'Variation and Gender'. In: *The Handbook of Language, Gender, and Sexuality: Second Edition*. Hoboken, US: John Wiley & Sons, Inc, pp. 85–102.

Meyerhoff, M. and S. Ehrlich (2019). 'Language, Gender, and Sexuality'. In: *Annual Review of Linguistics*. Vol. 5. 1. Annual Reviews, pp. 455–475.

Meyerhoff, M. and J. N. Stanford (2015). 'Tings change, all tings change: the changing face of sociolinguistics with a global perspective'. In: *Globalising sociolinguistics : challenging and expanding theory*. Ed. by D. Smakman and P. Heinrich. 1st ed. London: Routledge. Chap. 1, pp. 1–15.

Mondada, L. (2009). 'The Embodied and Negotiated Production of Assessments in Instructed Actions'. In: *Research on Language & Social Interaction* 42.4, pp. 329–361.

Moore, E. (2004). 'Sociolinguistic Style: A Multidimensional Resource for Shared Identity Creation'. In: *The Canadian Journal of Linguistics / La revue canadienne de linguistique* 49.3, pp. 375–396.

Moore, E. and R. Podesva (2009). 'Style, Indexicality, and the Social Meaning of Tag Questions'. In: *Language in Society* 38.4, pp. 447–485.

Mouatassim, Y. and E. H. Ezzahid (2012). 'Poisson regression and Zero-inflated Poisson regression: application to private health insurance data'. In: *European Actuarial Journal* 2.2, pp. 187–204.

Murphy, B. (2012). 'Exploring response tokens in Irish English a multidisciplinary approach: Integrating variational pragmatics, sociolinguistics and corpus linguistics'. In: *International Journal of Corpus Linguistics* 17.3, pp. 325–348.

Namba, A. (2011). 'Listenership in Japanese interaction: the contributions of laughter'. PhD thesis. University of Edinburgh / University of Tokyo.

Nilsson, J. (2015). 'Dialect accommodation in interaction: Explaining dialect change and stability'. In: *Language & Communication* 41, pp. 6–16.

Nishizaka, A. (2015). 'Facts and Normative Connections: Two Different Worldviews'. In: *Research on Language and Social Interaction* 48.1, pp. 26–31.

Norrick, N. R. (2012). 'Listening practices in English conversation: The responses responses elicit'. In: *Journal of Pragmatics* 44.5, pp. 566–576.

Ochs, E., E. A. Schegloff, and S. A. Thompson, eds. (1996). *Interaction and grammar*. Cambridge: Cambridge University Press.

Ogden, R. (2006). 'Phonetics and social action in agreements and disagreements'. In: *Journal of Pragmatics* 38.10, pp. 1752–1775.

— (2012). 'The Phonetics of Talk in Interaction - Introduction to the Special Issue'. In: *Language and Speech* 55.1, pp. 3–11.

— (2017). *An Introduction to English Phonetics*. 2nd ed. Edinburgh: Edinburgh University Press, p. 240.

O'Keeffe, A. and S. Adolphs (2008). 'Response tokens in British and Irish discourse: Corpus, context and variational pragmatics'. In: *Variational Pragmatics*. Ed. by K. P. Schneider and A. Barron. Amsterdam: John Benjamins, pp. 69 –98.

Oreström, B. (1983). *Turn-taking in English conversation*. Lund: Gleerup.

Pichler, H. (2010). 'Methods in discourse variation analysis: Reflections on the way forward'. In: *Journal of Sociolinguistics* 14.5, pp. 581–608.

— (2013a). 'Data, methodology, and theoretical framework'. In: *The Structure of Discourse-Pragmatic Variation*. Ed. by H. Pichler. Vol. 13. John Benjamins. Chap. 2, pp. 21–49.

— (2013b). *The Structure of Discourse-Pragmatic Variation*. Vol. 13. Studies in Language Variation. Amsterdam: John Benjamins Publishing Company.

— (2016a). 'Introduction: discourse-pragmatic variation and change'. In: *Discourse-Pragmatic Variation and Change in English: New Methods and Insights*. Ed. by H. Pichler. Cambridge: Cambridge University Press, pp. 1–18.

— (2016b). 'Uncovering discourse-pragmatic innovations: innit in Multicultural London English'. In: *Discourse-Pragmatic Variation and Change in English: New Methods and Insights*. Ed. by H. Pichler. Cambridge: Cambridge University Press, pp. 59–85.

Podesva, R. (2007). 'Phonation type as a stylistic variable: The use of falsetto in constructing a persona'. English. In: *Journal of Sociolinguistics* 11.4, pp. 478–504.

Pomerantz, A. (1984). 'Agreeing and disagreeing with assessments: some features of preferred/ dispreferred turn shapes'. In: *Structures of Social Action*. Ed. by J. M. Atkinson and J. Heritage. Cambridge: Cambridge University Press, pp. 57–101.

R Core Team (2016). *R: A Language and Environment for Statistical Computing*. Vienna, Austria.

Raphael, L. J. (1972). 'Preceding Vowel Duration as a Cue to the Perception of the Voicing Characteristic of WordFinal Consonants in American English'. In: *The Journal of the Acoustical Society of America* 51.4B, pp. 1296–1303.

Raymond, C. W. (2018). 'On the relevance of dialect: Conversation analysis and contact linguistics'. In: *Journal of Sociolinguistics* 22.2, pp. 161–189. arXiv: `arXiv: 1011.1669v3`.

Reid, J. (1995). 'A study of gender differences in minimal responses'. In: *Journal of Pragmatics* 24.5, pp. 489–512.

Rigby, R. A., D. M. Stasinopoulos, and C. Akantziliotou (2008). 'A framework for modelling overdispersed count data, including the Poisson-shifted generalized inverse Gaussian distribution'. In: *Computational Statistics and Data Analysis* 53.2, pp. 381–393.

Robinson, J. D. (2012). 'Overall Structural Organization'. In: *The Handbook of Conversation Analysis*. Chichester, UK: John Wiley & Sons, Ltd, pp. 257–280.

Roger, D. and W. Nesshoever (1987). 'Individual differences in dyadic conversational strategies: A further study'. In: *British Journal of Social Psychology* 26.3, pp. 247–255.

Rosenfelder, I., J. Fruehwald, K. Evanini, S. Seyfarth, K. Gorman, H. Prichard, and J. Yuan (2014). *FAVE (Forced Alignment and Vowel Extraction) Program Suite v1.2.2 10.5281/zenodo.22281*.

RStudio, T. (2015). *RStudio: Integrated Development Environment for R*. Boston, MA.

Ruiter, J. P. de and S. Albert (2017). 'An Appeal for a Methodological Fusion of Conversation Analysis and Experimental Psychology'. In: *Research on Language and Social Interaction* 50.1, pp. 90–107.

Sacks, H. (1995). *Lectures on Conversation*. Vol. 1-2. Oxford, UK: Wiley-Blackwell, pp. 1–818. arXiv: `arXiv:1011.1669v3`.

Sacks, H., E. A. Schegloff, and G. Jefferson (1974). 'A simplest systematics for the organization of turn-taking for conversation'. In: *Language* 50, pp. 696–735.

Sankoff, G. (1973). 'Above and beyond phonology in variable rules'. In: *New Ways of Analyzing Variation in English*. Ed. by C. J. N. Bailey and R. W. Shuy. Washington: Georgetown University Press, pp. 107–118.

Saucier, G. (1994). 'Mini-Markers: A Brief Version of Goldberg's Unipolar Big-Five Markers'. In: *Journal of Personality Assessment* 63.3, pp. 506–516.

Schegloff, E. A. (1982). 'Discourse as an interactional achievement: Some uses of uh huh'and other things that come between sentences'. In: *Analyzing discourse: Text and talk*, pp. 71–93.

— (1993). 'Reflections on Quantification in the Study of Conversation'. In: *Research on Language & Social Interaction* 26.1, pp. 99–128.

— (1996). 'Turn organization: one intersection of grammar and interaction'. In: *Interaction and grammar*. Ed. by E. Ochs, E. A. Schegloff, and S. A. Thompson. Cambridge: Cambridge University Press, pp. 52–133.

— (2007). *Sequence organization in interaction: A primer in conversation analysis I*. Cambridge: Cambridge University Press, pp. 1–300.

— (2015). 'Conversational Interaction The Embodiment of Human Sociality'. In: *The Handbook of Discourse Analysis*. Hoboken, NJ, USA: John Wiley & Sons, Inc., pp. 346–366.

Schegloff, E. A. and H. Sacks (1973). 'Opening up Closings'. In: *Semiotica* 8.4, pp. 289–327.

Scheier, M. F. and C. S. Carver (1985). 'The Self-Consciousness Scale: A Revised Version for Use with General Populations'. In: *Journal of Applied Social Psychology* 15.8, pp. 687–699.

Schiffrin, D. (1987a). *Discourse markers*. Cambridge University Press, p. 364.

Schiffrin, D. (1987b). 'Discovering the Context of an Utterance'. In: *Linguistics* 25.1, pp. 11–32.

Schweitzer, A. and N. Lewandowski (2012). *Accommodation of backchannels in spontaneous speech*.

Selting, M. (2007). 'Lists as embedded structures and the prosody of list construction as an interactional resource'. In: *Journal of Pragmatics* 39.3, pp. 483–526.

Selting, M. and E. Couper-Kuhlen (2004). *Studies in Interactional Linguistics*. Ed. by S. Thompson and P. Hopper. Vol. 10. Amsterdam/Philadelphia: John Benjamins.

Sidnell, J. (2012). 'Basic Conversation Analytic Methods'. In: *The Handbook of Conversation Analysis*. Ed. by T. Stivers and J. Sidnell. Chichester, UK: John Wiley & Sons, Ltd, pp. 77–99.

Sidnell, J. and T. Stivers (2012). *The Handbook of Conversation Analysis*. Ed. by J. Sidnell and T. Stivers. New York, UNKNOWN: John Wiley & Sons, Incorporated.

Silge, J. and D. Robinson (2017). *Text Mining with R*.

Sloetjes, H. and P. Wittenburg (2008). *Annotation by category - ELAN and ISO DCR*. Marrakech.

Sorjonen, M.-L. (2001). *Responding in Conversation. A study of response particles in Finnish*. John Benjamins.

Steensig, J. and T. Heinemann (2015). 'Opening Up Codings?' In: *Research on Language and Social Interaction* 48.1, pp. 20–25.

Stivers, T. (2004). 'No no no and Other Types of Multiple Sayings in Social Interaction'. In: *Human Communication Research* 30.2, pp. 260–293.

— (2008). 'Stance, Alignment, and Affiliation During Storytelling: When Nodding Is a Token of Affiliation'. In: *Research on Language and Social Interaction* 41.1, pp. 31–57.

— (2012). 'Sequence Organization'. In: *The Handbook of Conversation Analysis*. Chichester, UK: John Wiley & Sons, Ltd, pp. 191–209.

— (2015). 'Coding Social Interaction: A Heretical Approach in Conversation Analysis?' In: *Research on Language and Social Interaction* 48.1, pp. 1–19.

Stubbe, M. (1998). 'Are you listening? Cultural influences on the use of supportive verbal feedback in conversation'. In: *Journal of Pragmatics* 29.3, pp. 257–289.

Szczepek Reed, B. (2007). *Prosodic Orientation in English Conversation*. Hampshire: Palgrave Macmillan.

— (2009). 'Prosodic orientation: A practice for sequence organization in broadcast telephone openings'. In: *Journal of Pragmatics* 41.6, pp. 1223–1247.

— (2010). 'Speech rhythm across turn transitions in cross-cultural talk-in-interaction'. In: *Journal of Pragmatics* 42.4, pp. 1037–1059.

— (2011). 'Beyond the Particular: Prosody and the Coordination of Actions'. In: *Language and Speech* 55.1, pp. 13–34.

Szczepek Reed, B. (2014). 'Phonetic practices for action formation: Glottalization versus linking of TCU-initial vowels in German'. In: *Journal of Pragmatics* 62, pp. 13–29.

Tagliamonte, S. A. (2006). *Analysing Sociolinguistic Variation*. Cambridge: Cambridge University Press, p. 298. arXiv: `arXiv:1011.1669v3`.

— (2016). 'Antecedents of innovation: exploring general extenders in conservative dialects'. In: *Discourse-Pragmatic Variation and Change in English: New Methods and Insights*. Ed. by H. Pichler. Cambridge: Cambridge University Press, pp. 115–138.

Tamminga, M., L. MacKenzie, and D. Embick (2016). 'The dynamics of variation in individuals'. In: *Linguistic Variation* 16.2, pp. 300–336.

Thompson, S. A., B. A. Fox, and E. Couper-Kuhlen (2015). *Grammar in Everyday Talk*. Cambridge: Cambridge University Press.

Tolins, J. and J. E. Fox Tree (2014). 'Addressee backchannels steer narrative development'. In: *Journal of Pragmatics* 70, pp. 152–164.

Tottie, G. (1991). 'Conversational Style in British and American English: The case of backchannels'. In: *English Corpus Linguistics1*. Ed. by K. Aijmer and B. Altenberg. London and New York: Longman. Chap. 17, pp. 254–271.

Tracy, K. (1993). 'It's an Interesting Article!' In: *Research on Language & Social Interaction* 26.2, pp. 195–201.

Ward, N. (2006). 'Non-lexical conversational sounds in American English'. In: *Pragmatics {&} Cognition* 14.1, pp. 129–182.

Waters, C. (2016). 'Practical strategies for elucidating discourse-pragmatic variation'. In: *Discourse-Pragmatic Variation and Change in English: New Methods and Insights*. Ed. by H. Pichler. Cambridge: Cambridge University Press, pp. 41–56.

Watts, R. J. (1989). 'Taking the pitcher to the well': Native speakers' perception of their use of discourse markers in conversation'. In: *Journal of Pragmatics* 13.2, pp. 203–237.

White, S. (1989). 'Backchannels across Cultures: A Study of Americans and Japanese'. In: *Language in Society* 18.1, pp. 59–76.

Wieder, D. L. (1993). 'On the compound questions raised by attempts to quantify conversation analysis' phenomena part 2: the issue of incommensurability'. In: *Research on Language & Social Interaction* 26.2, pp. 213–226.

Wilkinson, S. and C. Kitzinger (2014). 'Conversation Analysis in Language and Gender Studies'. In: *The Handbook of Language, Gender, and Sexuality: Second Edition*. Hoboken, US: John Wiley & Sons, Inc, pp. 141–160.

Williams, G. and A. Mahmoud (2017). 'Mining Twitter Feeds for Software User Requirements'. In: *Proceedings - 2017 IEEE 25th International Requirements Engineering Conference, RE 2017*. IEEE, pp. 1–10.

Williams, L. (1977). 'The perception of stop consonant voicing by Spanish-English bilinguals'. In: *Perception & Psychophysics* 21.4, pp. 289–297.

Winter, B. and M. Wieling (2016). 'How to analyze linguistic change using mixed models, Growth Curve Analysis and Generalized Additive Modeling'. In: *Journal of Language Evolution* 1.1, pp. 7–18.

Wong, D. and H. Kruger (2018). 'That's Right A Multifactorial Analysis of the Selection of Backchannel Structures in British English'. In: *Corpus Approaches to Contemporary British Speech: Sociolinguistic Studies of the Spoken BNC2014*. Ed. by V. Brezina, R. Love, and K. Aijmer. Routledge. Chap. 8, p. 272.

Yaeger-Dror, M. and L. Hall-Lew (2002). 'Presidential Use of Negation'. In: *Proceedings of the Tenth . . .* 45, pp. 187–194.

Yngve, V. H. (1970). 'On getting a word in edgewise'. In: *Chicago Linguistics Society, 6th Meeting*, pp. 567–578.

Zama, A. and J. D. Robinson (2016). 'A Relevance Rule Organizing Responsive Behavior During One Type of Institutional Extended Telling'. In: *Research on Language and Social Interaction* 49.3, pp. 220–237.

Zellers, M. and R. Ogden (2013). 'Exploring Interactional Features with Prosodic Patterns'. In: *Language and Speech* 57.3, pp. 285–309.

Zimmerman, D. H. (1993). 'Acknowledgment Tokens and Speakership Incipiency Revisited'. In: *Research on Language and Social Interaction* 26.2, pp. 179–194.

Glossary

CGM Constant Glucose Monitoring. A filament is inserted subcutaneously, usually in upper arm or abdomen, and measures the glucose levels in the interstitial fluid. Interstitial glucose is very similar to blood glucose, but the reading lags 5 to 15 minutes behind blood glucose readings. Both CGM and FGM give trend arrows in addition to current glucose readings (stable, rising, falling). CGM systems transmit the glucose reading to a receiver every few minutes and alarm if glucose levels rise or fall above or below a certain threshold, or if glucose levels are changing quickly.. 113, 116, 125

DAFNE Dose Adjustment For Normal Eating, a diabetes education course run in the UK in which people with T1D are taught to carb-count and adjust their mealtime and background insulin according to their activities and what they eat.. 85, 88, 89, 113, 117

Dawn Phenomenon Rise of glucose levels in the early morning hours caused by hormones. Called Dawn Phenomenon because this rise happens at dawn as the body is preparing to wake up.. 118, 139

Dexcom Producer and name of the most popular CGM at the time the recordings were made. Sensors officially last 7 days but can be restartet several times until the results stop being reliable. The sensor sends readings to a reader, an animas insulin pump, or a smart device (phone, watch) via Bluetooth and it is possible to set alarms for high/low glucose readings.. 113, *see* CGM

Enlite CGM which communicates with Medtronic insulin pumps, offered by the company Medtronic.. 125, *see* CGM

Freestyle Libre Only FGM technology available in the UK at the time of recording. Sensors last 14 days and do not need to be calibrated by the user. Also called libre, flash, or freestyle by different participants.. 116, 117, 124, *see* FGM

glycaemic index The longer it takes the body to break down the carbohydrate contained in a meal, the lower that meal's glycaemic index is. Foods with a high glycaemic index like fizzy drinks, juice, fruit, granulated sugar, or white bread immediately raise the blood sugar and hence require a quick and high insulin response to keep blood glucose levels steady, while foods with a low glycaemic index like lentils or oats need cause a slower rise and hence need insulin over a longer period of time.. 134

hyperglycaemia High blood sugar levels (for a person with a functioning pancreas, those would be glucose readings above 8mmol, for people with diabetes above 10mmol).. 87

hypoglycaemia Low blood sugar levels (below 4mmol).. 139, 178

insulin pump Constant subcutaneous insulin delivery system as an alternative to mdi: an insulin pump constantly delivers insulin to the body, and the person using it can adjust the dosage simply by pressing a button instead of needing separate injections every time.. 113, 116, 125

ketogenic When less than 20-30 grams of carbohydrate per day are consumed, the body switches to using protein and fat as fuel. In this process, ketones are being produced as the body metabolizes the nutrients, and these ketones can be used as a source of energy. This metabolic state is called ketosis.. 135, 136, *see* ketosis

ketosis Metabolic state in which the body uses ketones as a source of energy, induced either by fasting or by a ketogenic diet.. *see* ketogenic

low carb A diet that is low in carbohydrates. Low carb tends to be used for diets with less than 100 grams of carbohydrate per day, very low carb for diets with less then 50 grams of carbohydrate per day, and ketogenic for diets with less 30 grams of carbohydrate per day.. 134, *see* ketogenic

Smart Guard The Medtronic Enlite sensor and the Medtronic 640g insulin pump communicate via bluetooth, and the pump suspends the basal insulin if glucose readings are dropping to prevent hypoglycaemia.. 126, 127

Appendices

Appendix A

Recruitment Sheet

THE UNIVERSITY of EDINBURGH
School of Philosophy, Psychology and Language Sciences

How much **Carb** in that?

Talk about Type 1
The highs and the lows

KEEP CALM and Carry

A Glucose Monitor, Test Strips, Alcohol Wipes, Lancets, Insulin, Syringes, Snacks and a Glucagon

Type 1 Diabetes Mama

Hello!

Wanted: people over the age of 18 living with type 1 diabetes or carers of children with type 1 diabetes.

Help us find out about how people talk about the **most common challenges** in daily life with type 1 diabetes, and which **solutions** you and your health care team have come up with, and **what still remains a challenge** and a mystery to you.

Participation in this study involves a short interview, filling in a questionnaire, and participating in three rounds of conversations about different areas of life that affect and are affected by having type 1 diabetes. **It won't take more than two hours** ☺

Location: University of Edinburgh, Central George Square Campus (EH8 9AD)
Duration: max. 2 hours
Compensation: you will be paid £20

You're in? Contact Mirjam Eiswirth at s1322502@sms.ed.ac.uk or 074-82257583. We'll then arrange a time and date for the interview and discussion sessions.

Thank you!

Omgosh, did they just say their child is high?

Okay okay.. So... If my blood sugar is high.. and I eat 60 carbs.. And I'm going to go for a run.. And I feel a cold coming on.. How much insulin do I take right now?

SOMEONE I L♥VE HAS DIABETES
Support Diabetes Awareness

Appendix B

Consent Form

Consent for Participation, Use of Audio-recording, and Data Storage

Study title:	Analysing conversations about living with type one diabetes: common problems and best practice solutions
Principal Investigators:	Dr. Lauren Hall-Lew, Dr. Joseph Gafaranga, Dr. Josef Fruehwald
Researcher collecting current data:	Mirjam Eiswirth

What is this document? This document explains what kind of study we're doing, what your rights are, and what will be done with your data. If there are any special benefits or risks, they will be explained here. By filling in, signing and dating this document, you will be agreeing to participate and to let us use your data in specific ways. Please read the information below, then turn to the next page, tick all boxes that apply, and, if you are happy to proceed, sign and date where indicated at the end of the form.

Nature of the study. You are about to participate in a study which involves recording your speech. The recordings will take place at the University of Edinburgh at a time of your convenience. You will be interviewed by a researcher, fill in a questionnaire, and talk to all three other participants about diabetes related topics. We are interested in the experience of living with diabetes or caring for someone living with diabetes

Your session should last for up to 2 hours. You will be given full instructions shortly and will be able to ask any questions you may have.

Compensation. You will be paid GBP 20 for your participation in this study.

Risks and benefits. There are no known risks to participation in this study. The results of this study are going to be put together in a report for Diabetes Scotland and the Scottish Parliament in order to help develop policy to adapt it to the needs of people living with type 1 diabetes.

Confidentiality. Unless you explicitly indicate otherwise on the attached consent form, your recording and your data will not be associated with your name or with any other personal details that might identify you.

Voluntary participation and right to withdraw. Your participation is voluntary, and you may withdraw from the study at any time and for any reason. If you withdraw from the study during or after data gathering, we will delete your data and there is no penalty or loss of benefits to which you are otherwise entitled.

Contact information. This research is being conducted by the above-listed researchers at the University of Edinburgh. The researchers can be contacted at 0131 651 1836 or meiswirt@ed.ac.uk for questions or to report a research-related problem. You may contact the Linguistics & English Language Ethics committee at 0131 651 5510 or lel.ethics@ed.ac.uk if you have questions or comments regarding your rights as a participant in the research.

If you have any questions about what you've just read, please feel free to ask now.

Thank you for your help!

Now please complete the consent form on the next page.

Consent for Participation, Use of Recorded Speech, and Data Storage

Study title:	Analysing conversations about living with type one diabetes: common problems and best practice solutions
Principal Investigator:	Dr. Lauren Hall-Lew, Dr. Joseph Gafaranga, Dr. Josef Fruehwald
Researcher collecting current data:	Mirjam Eiswirth

PLEASE MARK EITHER 'YES' OR 'NO' FOR <u>EVERY</u> STATEMENT BELOW:

	Yes	No
<u>Consent for participation</u>:		
I consent to having my speech recorded for the specific research project identified above. I have been given the opportunity to ask questions.	☐	☐
I understand that I have the right to terminate this session at any point. The recording of my speech will be deleted at that time / returned to me on request.	☐	☐
<u>Anonymity/identification</u>:	**Yes**	**No**
I agree that my name and personal details will not be made public under any circumstances.	☐	☐
<u>Researcher use of recordings</u>:	**Yes**	**No**
I agree that these recordings and any measurements drawn from them may be **kept permanently in Edinburgh University archives** and used **for the specific research project** which made them.	☐	☐
I agree that these recordings and any analyses drawn from them **may be used by the above-named researchers, as well as by other qualified researchers, for teaching or research purposes, and in professional presentations and publications**. In the case of voice recordings, I understand that **my voice might be recognizable to those who know me.**	☐	☐
<u>General public use</u>:	**Yes**	**No**
I agree that these recordings **may be made publicly available for general use**, e.g. used in radio or television broadcasts, or put on the world-wide web. In the case of voice recordings, I understand that **my voice might be recognizable to those who know me.**	☐	☐

Name:_____ Email:_____

Signature: _____ Date: _/d_____ /m_____ /y_____

Appendix C

Sociolinguistic questionnaire

To make sure all your data is anonymous, please invent a name you would like me to use when referring to you and your data in this research project.

_____

What is your year of birth?	What is your gender?	What is your ethnicity?
_____	_____	_____

Where did you grow up? (Be as specific as you like)

_____

Where did you go to secondary school?

_____

Where do you live now? (If in Edinburgh or Glasgow, which part of town?)

_____

Have you ever lived anywhere else? If yes, where and for how long?

_____

_____

What is your current job?

_____

What was your first job (and when)?

_____

_____

What is your highest degree?

_____

If you went/are going to University, what did/do you study?

_____

_____

What was your/ your child's latest HbA1c?   _____   prefer not to say

Appendix D

Topical Prompts

Diabetes Technology and Healthcare

Your health care team

Closed-loop system /
Artificial pancreas

Diabetes in you Daily Life

Diabetes and friends

A diabetic needs a best friend that they can list all the food they ate in a low and aren't judged

8

Diabetes and relationships

I'M SORRY FOR WHAT I SAID WHEN MY BLOOD SUGAR WAS LOW and high

9

"THINGS IN LIFE THAT CAN MAKE YOUR BLOOD SUGAR HIGH / LOW...

... VOLUME 1"

Believe it or not..

These are emergency gummy bears. No you can NOT have any!

10

Diabetes at work

Do you openly talk about your diabetes at work?
Yes - 65%
Somewhat - 25%
No - 10%
N=278

Do you keep back up insulin at work?
Always - 37%
Sometimes - 20%
Never - 43%
N=271

Have you taught any of your coworkers how to handle a diabetes emergency situation?
Yes 61%
No 39%
N=276

Have you ever brought up your diabetes during the interview process?
Yes 40%
No 60%
N=260

11

Diabetes and Mental Health

THE FIRST YEAR AFTER A DIABETES DIAGNOSIS IS LIKE...

I can't diabetes today.

Don't make me diabetes today....

type 1 diabetes memes

MINDFULNESS

Access to

psychological

support

Coping strategies
for days
when you feel like
diabetes
is holding you back

Source list

Image Sources in order of occurrence of pictures in pptx

- https://www.freestylelibre.co.uk/freestyle-libre-starter-pack-english.html (2016-12-05)
- http://www.medtronicdiabetes.com/res/img/misc/home-fam-revel.jpg (2016-12-05)
- http://www.shootuporputup.co.uk/wp-content/uploads/2016/04/P1018040.jpg (2016-12-05)
- http://insulinpens.com/wp-content/uploads/insulinpens.jpg (2016-12-05)
- http://diabetesneeds.co.uk/wp-content/uploads/2014/03/Education-Dafne.jpg (2016-12-05)
- http://www.berkshirewestdiabetes.org.uk/images/dafne-2# (2016-12-05)
- http://www.abta.org/assets/images/stock-photos/your-medical-team.jpg (2016-12-05)
- https://cdn.diabetesdaily.com/wp-content/blogs.dir/21/files/2016/10/artificial-pancreas-medtronic-fda-medtech.jpg (2016-12-05)
- http://kidshealth.org/EN/images/headers/K-diabetesSchool-enHD-AR1.gif (2016-12-05)
- https://i.ytimg.com/vi/7hjzAqfOqss/maxresdefault.jpg (2016-12-05)
- https://s-media-cache-ak0.pinimg.com/236x/a9/fb/32/a9fb329f54d9cf07925ca0ced6210879.jpg (2016-12-05)
- https://s-media-cache-ak0.pinimg.com/236x/4d/83/bc/4d83bc5bd57a40f4e5186729eddc5e6d.jpg (2016-12-05)
- http://ih1.redbubble.net/image.237254856.7009/raf,750x1000,075,t,101010:01c5ca27c6.jpg (2016-12-05)
- https://s-media-cache-ak0.pinimg.com/originals/3e/67/e6/3e67e6f58b4da526a87d1b4c705442a3.jpg (2016-12-05)
- https://s-media-cache-ak0.pinimg.com/736x/1a/28/48/1a2848be0849c39db3583cfaae9909e6.jpg (2016-12-05)
- https://myglu-content-production.s3.amazonaws.com/secure-wp-content-glu/uploads/2013/08/Work-QOTD-Infographic-v1.0.jpg (2016-12-05)
- https://peiyansesarma.files.wordpress.com/2016/04/the-first-year.jpg?w=500&h=316 (2016-12-05)
- https://yogafordiabetesblog.files.wordpress.com/2016/08/i-cant-diabetes-today.jpg?w=1200 (2016-12-05)
- http://www.haideemerritt.com/wp-content/uploads/2015/03/fingerprick_fire.jpg (2016-12-05)
- https://s-media-cache-ak0.pinimg.com/564x/a5/0b/3e/a50b3eb380140c6ba08e865fa21ffd79.jpg (2016-12-05)
- http://static.wixstatic.com/media/44a39b_208e2d9719484714a3061951cd3a9203.jpg (2016-12-05)

18

Appendix E

Interview Questions

The interviewer was given the following outline and interview questions for the semi-structured interviews:

Background information

This interview has three broad topics: Diagnosis, Support, and Impact. In the diagnosis section, I want to know about how and when the participant was diagnosed, and what happened immediately after diagnosis did they get their diabetes education in hospital, who helped them and how, etc.?

In the section on support I want to know about their experience with the NHS, regular check-up meetings, but also about support from their family and friends, and where they get information about diabetes. The support-section also includes a question on involvement with local support groups and volunteering.

The third section on impact and change asks about (a) how diabetes has changed the participant's life, and (b) what kind of change and development the participant would like to see with respect to the treatment of diabetes.

Interview Questions

Diagnosis

- When and how were you/your child diagnosed?

- What kinds of symptoms did you notice?

- Does diabetes run in your family?

- Did you know what was going on and how to deal with it?

- What happened after diagnosis?

Support

- Where do you get your information about diabetes and treatment options?

- What are the most valuable channels for you?

- Are you involved in any diabetes support groups or have you attended any events organised by Diabetes Scotland/Diabetes UK? What is your experience with them?

- How often do you see a doctor/nurse, and how do your meetings usually go?

- How do you feel about them, is there anything you'd change if you could?

- Have you ever been in hospital since your diagnosis? Why? How did it go?

Change and Impact

- How has diabetes changed your life? Give an example of a positive change and of a negative change.

- Is there something you wish existed that would make your life with diabetes easier?

- Where is improvement needed most urgently and why?

Appendix F

Debrief Sheet

Thank you for participating in our study on experiences of living with type 1 diabetes or caring for someone living with type 1 diabetes.

We are not only interested in the content of your conversations, but also in how you talked to each other. This research project looks at how people change the way they speak based on who they are talking to and how they are talking to each other do you sound more similar to your interlocutor when you agree? Do you use the same words or sentence structures? Do you sound more different when you disagree?

This is the linguistic side of this research project, which we could not tell you about in advance in order to avoid influencing how you interact with each other and the interviewer. The project consists of both parts the content analysis in order to provide a report for policy advice to Diabetes Scotland and the Scottish Parliament, and the linguistic analysis to answer questions about accommodation in interaction.

If you prefer your data not be used for linguistic research, you can ask for the recording to be deleted any time. If you do not want your data to be used for this study, please contact either the Linguistics & English Language Ethics Committee at 0131 651 5510 or lel.ethics [at] ed.ac.uk or Dr Lauren Hall-Lew at 0131 651 1836 or Lauren.Hall-Lew [at] ed.ac.uk or the researcher at meiswirt [at] ed.ac.uk.

If you wish to amend your consent form now that you know about the content aspect as well as the linguistic side of this research project, you can do so now or any time later by contacting any of the parties named above.

If you have any questions about either part of the analysis, feel free to ask any member of the research team.

Thank you for your participation.

Appendix G

Extracts for IRR

Date	Speakers	Topic	Start	End	(n) Tokens	– Laugh
20161217	Chris, Emma	Healthcare	03:24	05:10	9	9
20161217	Chris, Lavina	Mental Health	17:12	19:40	8	8
20161217	Chris, Samantha	Social Life	06:55	09:10	26	25
20161217	Emma, Samantha	Mental Health	09:30	12:00	32	31
20161217	Lavina, Emma	Social Life	07:55	10:20	22	20
20161217	Lavina, Samantha	Healthcare	00:38	04:00	18	17
20170117	Angie, DaisyRae	Healthcare	11:27	14:00	42	41
20170117	Angie, Darren	Social Life	03:00	05:24	28	27
20170117	DaisyRae, Darren	Mental Health	16:15	18:10	24	24
20170117	Rose, Angie	Metal Health	01:36	03:50	25	22
20170117	Rose, DaisyRae	Social Life	14:17	end	10	8
20170117	Rose, Darren	Healthcare	21:00	24:14	19	18
20170126	Lily, Tess	Social Life	09:48	12:56	36	26

Date	Speakers	Topic	Start	End	(n) Tokens	– Laugh
20170126	Lily, Tomek	Mental Health	13:08	17:04	30	30
20170126	Lily, Velomi-nati	Healthcare	05:52	08:34	15	15
20170126	Tess, Velomi-nati	Mental Health	23:17	26:17	23	21
20170126	Tomek, Tess	Healthcare	19:34	22:11	25	22
20170209	Tomek, Velominati	Social Life	25:36	29:36	18	18
20170209	Donna, Con-nor	Mental Health	04:50	06:32	15	13
20170209	Kinder-Surprise, Connor	Social Life	18:02	20:56	25	24
20170209	Kinder-Surprise, Donna	Healthcare	14:47	16:44	9	5
20170209	PuzzleB, Connor	Healthcare	16:19	21:52	27	23
20170209	PuzzleB, Donna	Social Life	12:00	14:30	25	21
20170209	PuzzleB, KinderSur-prise	Mental Health	22:17	25:17	36	33
Total					547	501

Appendix H

Coding Manual

This coding manual was written on 2 August 2018. It is to be used in conjunction with the interactional analysis chapter in this thesis (chapter 5). The section briefly summarising the action types in the present manual is superseded by the most recent version of the interactional analysis chapter. The second coder was given the draft version from August 2018 and directed to the tabular overview at the end to guide his coding decisions.

Definition of Listener Responses

In dyadic interactions, participants take so called turns at talk, with one speaker holding the floor and the other listening. Listening is by no means a silent or passive activity. On the contrary, listeners produce a variety of responses that in turn influence how the speaker's talk develops.

We define as listener responses all those things the person who is **not** holding the floor at a given point in time utters while the other person continues to talk. These responses can stand in overlap with the ongoing talk, or come in brief pauses within this longer stretch of talk. Very often they are brief vocalisations like *mm, mm hm, uh-huh, okay, right, yeah, yes*, but they can also be longer utterances like *That's great!* or *Oh my gosh!*, word-supplies if the speaker indicates word-search trouble, brief questions to get the speaker to elaborate on something, or voicing what a character in the speaker's story might have said. What makes these diverse actions or combinations of lexical material listener responses is that they respond to an ongoing stretch of talk by the current main speaker and make no attempt to interrupt or end it.

Multiple sayings and longer listener responses

Speakers may produce several listener response tokens in a row. Sometimes these are multiple sayings of the same token, sometimes participants produce different responses immediately adjacent to each other. Principled decisions need to be made on how to distinguish what counts as one listener response. Following Stivers (2004) I treat repeated items as doing a single action if they are under one intonation contour, form one (listener-)TCU, and are oriented to as one single action by the continuing speaker.

Repeated sayings are different from false starts - instances of false starts and self-repair are marked by hesitations, pauses, and separate intonation contours. Only the final production is commonly counted in sociolinguistic coding (Tagliamonte, 2006: p.94), which means false starts are counted as one token together with the final formulation. Given that false starts in listener responses are extremely rare, they will not be coded as a separate category.

Complex listener responses that contain several different actions are coded as those different actions, even if all the lexical material is under one intonation contour.

Action Types

As mentioned above, listeners can do a number of different actions with their responses. What a listener response is doing can be gleaned from the sequential context in which it occurs - i.e. what kind of speaker-talk comes before it, and what kind of speaker-talk follows it. The action types are as follows:

1. acknowledgements

2. markers of surprise

3. first assessments

4. second assessments

5. self-initiated other-repair (Word-search completions)

6. other-initiated self-repair (questions, corrections)

7. joint utterances (also referred to as 'collaborative completions')

Examples for all action types are contained in the CA chapter explaining the coding scheme in more depth. Below is a schematic sequential pattern for each of those action types. 1. is always the part of the ongoing talk preceding the listener

response, 2 the listener response, and 3. the speaker's reaction. In some cases, the sequence can be slightly longer because a clarification or ratification is needed.

Acknowledgements

1. **Speaker:** ongoing multi-unit-turn

2. **Listener:** acknowledgement

3. **Speaker:** continuation of ongoing multi-unit-turn

Markers of surprise

1. **Speaker:** ongoing multi-unit-turn (containing something surprising, unusual, or unexpected)

2. **Listener:** surprise mark

3. **Speaker:** continuation of ongoing multi-unit-turn (often elaboration of the thing marked as surprising)

First assessments

1. **Speaker:** assessable

2. **Listener:** first assessment

3. **Speaker:** continuation

The format of the first assessment as a listener response is recognisable as a conventional assessment format:

[Response cry] +	[NP] +	[BE] +	[EVALUATOR]
oh	∅	∅	brilliant
∅	that	's	good
∅	that	's	great

TABLE H.1
SUMMARY PATTERN FIRST ASSESSMENT

Second assessments

1. **Speaker:** first assessment (potentially implicit)

2. **Listener:** second assessment (always preferred/aligning)

3. **Speaker:** (receipt+) continuation

Self-initiated other-repair

1. **Speaker**: repair initiation (through indication of trouble, i.e. hesitations, pauses, or false starts)

2. **Listener**: repair (word supply)

3. **Speaker**: (ratification; only if repair initiation was strongly marked) continuation

4. **Listener**: acknowledgement in overlap with continuation

Other-initiated self-repair

1. **Speaker:** repairable

2. **Listener:** repair-initiation, orienting to preceding talk as

   (a) surprising or unusual

   (b) incomplete

   (c) problematic in some way (issue with hearing or understanding)

3. **Speaker:** (repair+) continuation

Joint utterances

1. **Speaker:** highly projectable turn OR character in story speaks, response is relevant and projectable

2. **Listener:** collaborative completion or voicing

3. **Speaker:** continuation OR reaction to listener's voicing or suggestion

4. **Listener:** acknowledgement

5. **Speaker:** continuation

Annotation in ELAN - the technical side of things

The interactions have all been transcribed orthographically and are time-aligned. There are three tiers for each speaker:

1. The highest level in the hierarchy is the talk-tier, on which everything the speaker says is transcribed.

2. The second level is a 'listener response tier' for that speaker. This contains time-aligned annotations, in which the (non) lexical material of a given listener response is written down again.

3. The third level is the action type annotation. This level has a controlled vocabulary, so that for each annotated listener response the coder only needs to choose which of the seven actions described above is being done.

Steps in the Coding Process

1. I pseudo-randomly select an extract of each conversation that is 10 % of said conversation's length.

2. I share the training manual, my chapter draft with the action type analysis, the sound files, and ELAN files with empty annotation tiers with Zac

3. Zac identifies and annotates listener responses and does the action type coding

4. We meet to compare coding decisions and discuss problematic cases

5. I do the different IRR analyses

Appendix I

Overall Data Overview

Recording Date	Speaker 1 (S1)	Speaker 2 (S2)	Topic	Recording Duration	Words total	Listener Responses total	Words S1	Words S2	Listener Responses S1	Listener Responses S2
20161217	Chris	Emma	Health-care	17:16	2499	107	1292	1207	47	60
20161217	Chris	Lavina	Mental Health	23:37	3127	119	1214	1913	84	35
20161217	Chris	Samantha	Social Life	22:51	3551	173	2480	1071	26	147
20161217	Emma	Samantha	Mental Health	25:22	3899	204	1727	2172	108	96
20161217	Lavina	Emma	Social Life	24:19	3904	214	2182	1722	82	132
20161217	Lavina	Samantha	Health-care	21:48	3109	176	1746	1363	56	120
20170117	Angie	DaisyRae	Health-care	22:49	4477	331	2372	2105	109	222
20170117	Angie	Darren	Social Life	24:25	4408	262	2197	2211	210	52

Recording Date	Speaker 1 (S1)	Speaker 2 (S2)	Topic	Recording Duration	Words total	Listener Responses total	Words S1	Words S2	Listener Responses S1	Listener Responses S2
20170117	DaisyRae	Darren	Mental Health	19:06	3054	194	1474	1580	59	35
20170117	Rose	Angie	Metal Health	20:49	4014	207	1730	2284	106	101
20170117	Rose	DaisyRae	Social Life	15:58	2017	149	1073	1944	91	58
20170117	Rose	Darren	Health-care	32:33	5296	166	2403	2893	122	44
20170126	Lily	Tess	Social Life	29:51	7136	299	2578	4558	172	127
20170126	Lily	Tomek	Mental Health	27:05	4739	210	2392	2437	120	90
20170126	Lily	Velominati	Health-care	27:31	4587	230	2516	2071	133	97
20170126	Tess	Velominati	Mental Health	28:50	6353	228	3913	2440	136	92

Recording Date	Speaker 1 (S1)	Speaker 2 (S2)	Topic	Recording Duration	Words total	Listener Responses total	Words S1	Words S2	Listener Responses S1	Listener Responses S2
20170126	Tomek	Tess	Health-care	27:01	6362	279	2212	4150	156	123
20170209	Tomek	Velominati	Social Life	40:49	7493	254	3950	3543	130	124
20170209	Donna	Connor	Mental Health	16:48	3407	122	909	2498	71	51
20170209	KinderSur-prise	Connor	Social Life	29:44	6095	301	2693	3402	112	189
20170209	KinderSur-prise	Donna	Health-care	19:05	3589	135	2529	1060	40	95
20170209	PuzzleB	Connor	Health-care	27:35	6483	266	3419	3064	129	137
20170209	PuzzleB	Donna	Social Life	25:16	5549	267	3370	2179	152	115
20170209	PuzzleB	KinderSur-prise	Mental Health	30:11	6706	307	3819	2887	168	139

263