Journal of Pragmatics 170 (2020) 172–188

Contents lists available at ScienceDirect

Journal of Pragmatics

journal homepage: www.elsevier.com/locate/pragma

Increasing interactional accountability in the quantitative analysis of sociolinguistic variation

Mirjam Elisabeth Eiswirth

University of Duisburg-Essen, Sociolinguistics Lab, Department of Anglophone Studies, Universitätsstr. 12, 45141 Essen, Germany

A R T I C L E I N F O

Article history:
Received 22 January 2020
Received in revised form 17 August 2020
Accepted 19 August 2020
Available online 15 September 2020

Keywords:
Backchannels
Sociolinguistic variable
Interaction/conversation analysis
Language
Variation
Mixed methods

A B S T R A C T

Language is an interactional phenomenon, yet most studies of language variation focus on the community's or the individual speaker's production more than the interactional context. This paper argues that there are some cases in which we need to situate sociolinguistic variation in the interactional space and structure. It highlights the theoretical motivation for integrating interactional and variationist approaches, demonstrates how this can be done methodologically, and discusses the theoretical and methodological impact of such an approach on studies of language variation more broadly.

This theoretical contribution is made based on the example of Listener Responses, small vocalised feedback given during ongoing interlocutor-talk. They exemplify key issues that result from not rooting variables above the level of the phoneme in interactional structure: the variable's definition, delimitation, and quantification have a low level of emic validity relative to what actually happened in the interaction, and the analysis risks conflating social with interactional factors.

I introduce and argue for an interactionally rooted definition and quantification of Listener Responses as couched in the Speaker's ongoing (multi-unit) turn, and quantified relative to the length of this turn.

1. Introduction

Language is usually produced in interaction, but studies of language variation tend to analyse it relative to the wider community, or the individual(s) involved. However, variables above the level of the phoneme – for example, Listener Responses – are rooted in the interaction and need to be defined and quantified with this in mind, while the interactional structure is not a definitional feature of the phonological variable.[1] Listener Responses are frequently called backchannels (Kogure, 2003; Oreström, 1983; White, 1989; Wong and Kruger, 2018; Yngve, 1970) and refer to the small things Listeners do to support the Speaker's ongoing talk. Extract 1 briefly illustrates some Listener Responses. Angie describes a technical device for her son's type 1 diabetes management, and DaisyRae repeatedly provides vocalised feedback: *oh brilliant* (line 4), or *yeah* (lines 7, 9). All Listener Responses are marked with arrows.

E-mail address: mirjam.eiswirth@uni.due.de.

[1] Of course both the interaction and the interlocutor impact on the realisation of such phonological variables, as has been shown in many studies of Communication Accommodation Theory and studies of phonetics of talk-in-interaction. However, in these cases the interaction and the interlocutor are considered additional conditioning effects, while they are part of the variable definition for phenomena like Listener Responses.

https://doi.org/10.1016/j.pragma.2020.08.018

(1) Angie and DaisyRae, SmartGuard, minutes 07:17–07:30.

```
1         Angie:      and then it’ll (.) it’s got this thing called
2                     smart guard which can um (.)-
3         Angie:      [!] suspend the basal rate?=
4    →    DaisyRae:   =°°oh brillia[nt°°-
5         Angie:                   [so [if it knows that he’s (.)
6                                          going to (.) 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 (.)=
                            [...]
```

Work on Listener Responses exemplifies key issues that result from not rooting variables above the level of the phoneme in interactional structure: the same phenomenon has been defined, delimited, and quantified in different ways in previous work, leading to low emic validity and low comparability across studies. Furthermore, analyses risk conflating social with interactional factors.

Previous work on variation in Listener Response frequency has looked at inter-group differences with respect to culture (Heinz, 2003; O'Keeffe and Adolphs, 2008; White, 1989) and gender (Fellegy, 1995; Kogure, 2003; Reid, 1995). While both vary to some extent in the corpus under analysis here, the sample is more balanced for gender, making this the more useful social variable to demonstrate what can be done with the proposed methodology.[2] Crucially, the contributions of this paper are theoretical and methodological and can be applied to any other social variable, for example age, ethnicity, or other more locally constructed categories. I introduce and argue for an interactionally rooted definition and quantification of Listener Responses: as couched in the Speaker's ongoing (multi-unit) turn, and quantified relative to the length of this turn.

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 the individual's linguistic production (as we would with phonological variables), but that of the interlocutor – in this case, the length of the turn that is being responded to. To make these points, I first introduce the discussion around definition(s) of the sociolinguistic variable above the level of the phoneme and then present background on Listener Responses as a case study. From there, I move on to an interactionally rooted definition of Listener Responses as the variable, present the data and quantitative methods, and summarise my results, before discussing their theoretical and methodological implications for our understanding of the sociolinguistic variable, work on Listener Responses, and other variables above the level of the phoneme.

2. Defining the sociolinguistic variable above the level of the phoneme

In his foundational sociophonetic studies in the 1960s, Labov summarised the ‘most useful properties of a linguistic variable’ as follows: it should be (1) frequent, (2) structural, or ‘integrated into a larger system of functional units’ (Labov, 1963, p. 279) as much as possible, (3) ‘highly stratified’ by social and/or linguistic factors, and (4) salient. This definition is particularly well suited to phonetic and phonological variables, and frames them as the ‘most useful’ variables, given that they have the ‘most useful properties’. Additionally, for this type of variable it is relatively straightforward to fulfil the principle of accountability (Labov, 1972), i.e. the need to account for all *possible* occurrences, including absences in places where it could have occurred from a structural point of view.

This is challenging above the level of the phoneme, and scholars have put forward notions of semantic, formal, functional, procedural, or derivational equivalence as definitional criteria (Buchstaller, 2008, 2006; Dines, 1980; Lavandera, 1978; Pichler, 2013, 2010; Sankoff, 1973; Terkourafi, 2011). However, the issue of fulfilling the principle of accountability remains notoriously challenging. It is only possible for some discourse-level variables to count all *possible* places of occurrence. Quotatives are one example. If they are defined functionally as ‘all strategies used to introduce reported speech, sounds, gesture and thought by self or other’ (Buchstaller, 2006, p. 5), the absence of quotatives in contexts where we might expect one can be counted.

However, a large number of discourse-pragmatic variables are highly difficult to define and delimit in this way, because their optionality might be part of their definition, as is the case for Listener Responses. Here, 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 consideration of potential structural constraints, which can lead to

[2] 11 participants are from Scotland, two from England, one from Ireland, one is Finnish-American, and one is originally from Poland. If we stratify the sample by gender, 11 participants identify as female, 5 as male.

a misinterpretation of interactionally driven variation as a 'social effect', as Murphy (2012) discusses with respect to Listener Responses and gender. Additionally, these frequency counts complicate inferential statistical analysis, so that we can usually only work with descriptive comparisons across groups, without statistically factoring out random effects of, for example, individual variability.

Neg-tags are similar in their optionality, but it is possible to describe a full set of formal variants based on derivational equivalence, and do a distributional analysis, as shown in Pichler (2013). Pichler draws on CA tools to develop a function-based coding scheme and includes function in a multivariate analysis. In a similar vein, (Barron et al. (2017), Barron and Schneider (2009), and Schneider and Barron (2008) have introduced the field of Variational Pragmatics (VP) to integrate the two perspectives and understand variation based on the interplay between different social factors and levels of pragmatic analysis. I build on the steps in Pichler (2013) towards integrating CA into the analysis of variation above the level of the phoneme. This also aligns with the broad vision of VP. The methodology proposed here goes beyond Pichler (2013) by basing not just the description of functional variants but also the definition of the variable itself on interactional structure, and by normalising frequency at a more fine-grained, turn-by-turn level. Further, Listener Response frequency is normalised relative to the number of words produced by the turn-holder (i.e. the *other* participant) as the interactionally relevant context here. This study contributes to VP by providing an analysis of a variable at the so far understudied organisational level (Barron et al., 2017, p. 101).

3. Listener Responses as a case study

Quantitative studies on Listener Responses showcase the theoretical and methodological issues outlined above. At the same time, we can build on four decades of qualitative work to address these challenges.

Broadly defined, Listener Responses are things Listeners do to support the ongoing talk. They were introduced as and are often called *backchannels* (Kogure, 2003; Oreström, 1983; White, 1989; Wong and Kruger, 2018; Yngve, 1970), but have also been referred to as *minimal responses* (Fellegy, 1995; Reid, 1995), *reactive tokens* (Clancy et al., 1996), or *response tokens* (O'Keeffe and Adolphs, 2008). They have been found to vary with respect to gender, L1, or culture, and discussed as a variable at the organisational level in VP (Barron et al. (2017, p. 94) with particular reference to McCarthy (2003) and Tottie (1991)). This terminological variation reflects two issues with quantitative work from a qualitative perspective: there is a range of definitions and delimitations of the variable, and approaches to quantification vary. Neither usually attend to interactional structure, which leads to two complications: first, the numbers being analysed do not reflect the interactional reality of the conversations and lose crucial detail; second, the lack of a structural definition makes more complex quantitative analysis difficult. Furthermore, comparing findings across studies becomes highly challenging, if not impossible.

Let us first look at those variable definitions, delimitations, and quantifications of Listener Responses in quantitative studies, and then explore how integrating a qualitative approach might address the challenges just mentioned.

3.1. Quantitative work on Listener Responses

Definitional criteria of Listener Responses as a variable have been based on form (particularly length), or a combination of form and placement. Some consider specific (though varying!) sets of lexical items (Fellegy, 1995; Gardner, 1998); others also include longer responses and multimodal cues like head nods, laughter, or eye blinks (Bavelas et al., 2000; Brunner, 1979; Kogure, 2003; Maynard, 1997). With respect to the combination of form and placement, sometimes only responses that are positioned *between*, that is, not in overlap with, parts of ongoing talk and do not exceed a certain length are considered (Schweitzer and Lewandowski, 2012); at other times overlap with ongoing talk is accepted as long as the response does not exceed a certain length (White, 1989).

The three most common approaches taken to quantifying the frequency of Listener Responses are (1) per million words in corpus, (2) per X minutes, or (3) relative to Speaker changes (Eiswirth 2020a). None of these is particularly sensitive to the interaction and the distribution of Speaker and Listener roles.

The first metric, per million words, is standard in corpus linguistic work (O'Keeffe and Adolphs, 2008; Murphy, 2012). Murphy (2012) finds that her male Irish English speakers produce more Listener Responses per million words in the corpus than her female Irish English participants. However, she demonstrates that what seems like a gender difference is in fact explained by an uneven distribution of interactional roles: in her recordings, men spend more time as Listeners and thus produce more responses relative to the total words in the corpus than women. Hence, a metric like 'per million words' does not give us information about participants' behaviour as Listeners. Instead, it reflects how much time they spend in the Speaker or the Listener role.

The second metric, counting the number of responses in a given number of minutes in the interaction (Bavelas et al., 2000; Maynard, 1990; Schweitzer and Lewandowski, 2012; Tottie, 1991), is prone to the same confound: if we do not take the Speaker and Listener role distribution into account, seemingly social differences might easily be skewed by interactional factors.

The third approach is to compare the number of Speaker changes to take over the floor to the number of Speaker changes to Listener Responses in percent (Brunner, 1979; Clancy et al., 1996; Kogure, 2003). 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 contributions into account. It could therefore still be skewed if one person took more and/or longer turns at talk than the other.

There is a fourth approach to quantifying frequency with its origins in social psychology: the number of Listener Responses by one participant in an interaction is divided by the total number of interlocutor words and then multiplied by 100 (mean turn length) (Duncan and Fiske, 1977). This approach acknowledges that Listeners respond to what the *other* person in the interaction does and thus avoids conflating the distribution of Listener- and Speaker-roles with Listener Response frequency (see Murphy (2012)), but remains at a relatively low level of granularity – the overall conversation. I build on this metric here (see section 3.2).

3.1.1. Listener Responses and gender

Despite this variation in precise definition, delimitation, and frequency operationalisation, most studies on gender and Listener Response frequency find no difference in how often male or female Listeners produce responses. Oreström (1983) and Fellegy (1995) both count up responses in a given amount of time in British and American English respectively, while Kogure (2003) looks at Japanese interactions with a 'percent of speaker changes' metric. In data from South African English, Dixon and Foster (1998) find no main effect of Listener gender based on the metric proposed by Duncan and Fiske (1977), but an interaction between Listener and Speaker gender: Male Listeners produce more responses when listening to a woman than women listening to women. Similarly, Reid (1995) finds that female turn-holders receive more responses than male Speakers, irrespective of Listener gender.

The only study which does find a gender difference in Listener Response frequency is Stubbe (1998) on Maori and Pakeha speakers of New Zealand English. However, she uses a metric that is conceptually similar to 'per million words' in corpus, which means this gender difference might be driven by an uneven distribution of speaker roles (see Murphy (2012)). Both Fellegy (1995) and Reid (1995) note the importance of contextual and interactional factors and interlocutor gender, but neither puts forward a methodological approach to tackling these questions.

To summarise, Listener Responses exemplify the theoretical and methodological issues related to extending the sociolinguistic variable beyond the level of the phoneme: there is a lack of a shared, structurally rooted definition and delimitation of the variable, which leads to a diversity of approaches to frequency operationalisation. This reduces the emic validity of these definitions, the interpretability of the results relative to actual interactional processes, the possibilities for statistical analysis, and makes cross-study comparisons problematic. In terms of variation in Listener Response frequency based on gender, most studies find no gender difference but note the importance of interactional roles and contextual factors.

3.2. Towards a qualitatively rooted definition and quantification

In the following, I propose an interactionally rooted definition, delimitation, and frequency operationalisation of Listener Responses. This first key contribution allows us to address the issues outlined so far.

Interaction is the natural ally when attempting a structural definition of Listener Responses. There is ongoing debate about CA and quantification, but scholars agree that in order to create accountable, interactionally rooted coding schemes and quantifications, we need to have a solid qualitative understanding of the potential variable (Heritage, 1999; Schegloff, 1993). Thanks to four decades of work on different aspects of Listenership, we have such a solid qualitative understanding, based on which we can define the numerator (i.e. all possible realisations of the variable) and the denominator (i.e. the envelope of variation, see Schegloff (1993)).

At the local level of discourse organisation, CA scholars analyse how people take turns at talk and how they take and cede the conversational floor. Usually, one person 'takes a turn', then the next, and so on, in relatively quick succession (Sacks et al., 1974). However, participants can also decide to give the floor to one person for a longer period of time, in which they produce a so-called 'multi-unit-turn', for example a story (Mandelbaum, 2012). In those cases, the roles of Listener and Speaker need to be reaffirmed from time to time to keep the conversation going. Thus, 'Listening' does not imply being passive or quiet at all. Rather, Listeners regularly produce signals of their continued Listenership (Schegloff, 1982). I call those signals Listener Responses here to highlight their crucial role in co-creating the interactional space. This term focuses on the fact that the Listener actively *does* something and thus co-creates the interaction with the Speaker. Terms like 'backchannel' or 'minimal response' understate the importance of the Listener.

Thus, from the point of view of interactional structure, Listener Responses are all those things Listeners do in the Speaker's ongoing longer stretch of talk, which does not lead to the Listener taking over the floor and becoming the next Speaker (Jefferson, 1984; Schegloff, 1982). At an abstract level, we can represent this as follows:

Structural definition of a Listener Response.

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

Note that this delimitation of the numerator (what counts as a Listener Response) is primarily defined based on sequential structure rather than form: if the Listener says or does something, and the Speaker continues their ongoing talk, the Listener's contribution has been treated as a Listener Response. This focus on 'What happens next?' as the basis for our coding decisions is called the next-turn proof-procedure in CA (Clift, 2016; Heritage, 1984; Sidnell, 2012). Its importance will become evident when considering extracts containing some of the over 5200 (potential) Listener Responses I have worked with for this study: Prototypical ones in extract 2, a longer and more involved one in extract 3, and two more in extract 4. These are an instance which, based on its form, looks like a Listener Response but is in fact a segway into taking over the floor, and finally an instance of a floor transition that could be mis-counted as a Listener Response if we only considered the immediately following talk.

This first extract contains a number of prototypical, brief Listener Responses. Angie, the mother of a 10-year old with Type 1 Diabetes, is talking about her son's insulin pump and the type of glucose monitoring system 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, mostly very minimal and quiet ones (marks very quiet speech) in overlap with Angie's talk, but also two slightly longer ones in lines 4 and 25.

(2) Angie and DaisyRae, SmartGuard, minutes 07:17–07:50.

```
    Angie:      and then it'll (.) it's got this thing called
                smart guard which can um (.)-
    Angie:      [!] suspend the basal rate?=
→   DaisyRae:   =°°oh brillia[nt°°-
    Angie:                   [so  [if it knows that he's (.)
                                       going to (.) go low?
→   DaisyRae:                     [°°yeah°°.
    Angie:      or if [he's trending low?
→   DaisyRae:         [°yeah°.
    Angie:      then it'll suspend-
    Angie:      (.4)
    Angie:      the rate I think maybe (.)=
    Angie:      =twe:nty minute[s or something before-
→   DaisyRae:                  [°okay°-
    Angie:      he actually hits hypo [so:-
→   DaisyRae:                         [°yeah°-
    Angie:      and it it catch- (.)
    Angie:      like we were really-
    Angie:      (.4)
    Angie:      not very sure about it,=
    Angie:      [well I think you know we'll still
→   DaisyRae:   [°yeah°.
    Angie:      give Lucozade or sort of if we see the
    Angie:      arrow's going down.
→   DaisyRae:   °°yeah°°=
    Angie:      =but it does catch most of [them.
→   DaisyRae:                              [°°that's great.°°
    Angie:      = yeah I think just if [he's had a lot of-
→   DaisyRae:                          [°°yeah°°.
    Angie:      if he's got a lot of active insulin
```

Angie keeps the floor throughout this segment, while DaisyRae ratifies her Speakership by providing regular small responses. Most of them only consist of one word, *yeah* or *okay* (lines 7, 9, 14, 16, 22, 25, and 29); *oh brilliant* (line 4) and *that's great* (line 27) are slightly longer.

Many responses are placed at points where the turn constructional unit (TCU), that is, one of the segments that together make up a multi-unit turn, could be complete. Those points of potential grammatical, syntactic, and prosodic completion are called Transition Relevance Places (TRPs, see Clancy et al. (1996) and Ford and Thompson (1996)). We can see those for example in line 3 where Angie has completed a syntactic unit (lines 1–3, 15, 20). In all three cases, DaisyRae immediately follows the potentially complete turn with a Listener Response (lines 4, 16, 22), signalling that she is still listening and not putting the role distribution into question. However, it is not the case that responses only occur at TRPs: *yeah* (line 9) overlaps with Angie's ongoing TCU ‘or if he's trending low’ after ‘if’, when the TCU is still projectably to be completed.

What all these Listener Responses have in common is that they do not disrupt the Speaker's ongoing (multi-unit) turn and reaffirm the Speaker-Listener role distribution; each response is followed by more talk from the main Speaker. Sequential structure is important as a definitional criterion because Listeners can do more than just brief, comparatively quiet responses and still remain the Listener, especially in more animated conversations. They can, for example, voice what a character in a story might have said. Wilkinson and Kitzinger (2014, p. 148) nicely demonstrate that in those cases both the Speaker's and the Listener's contribution to the turn are oriented to as the ‘Speaker's turn’. This is what we can see in extract 3. Lavina has just told Samantha about her DAFNE experience (DAFNE is a structured Diabetes education course) and how her diabetes team suggested she get up several times every night to check her blood sugar. Samantha then provides a Listener Response by voicing a candidate answer.

(3) 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
49			to.'
50		Lavina:	[!] [no I'd-
51	→	Samantha:	[((laughing))
52		Lavina:	[I- did do tha[t-
53	→	Samantha	[°°mhm°°-
			LAVINA CONTINUES

Lavina recounts the question the nurses asked her at the training course (line 46), and Samantha latches onto this quote and provides an answer Lavina might have given (line 48/49). This is followed by laughter rather than more speech on Samantha's part, indicating that she is not out to take over the floor. She thereby treats her utterance as part of Lavina's turn. Lavina (line 50/52) corrects Samantha's continuation, in partial overlap with the end of Samantha's contribution and the following laughter. Samantha produces an unambiguous, very brief and quiet Listener Response in overlap with the end of Lavina's continuation of her story (line 53), reaffirming her own place as the Listener. Lavina then continues the story, and Samantha continues to provide brief signals of her Listenership.

This example demonstrates the importance of the next-turn proof-procedure in determining what is and is not a Listener Response – the variable definition is based on what happens next sequentially: the main Speaker continues their talk, and the Listener does not make a claim to the floor.

The next extract further underscores the importance of this approach. In extract 4 we see how a lexical item that is often employed as a Listener Response is instead used to start a new turn and change the Listener-Speaker roles (lines 34, 51, 54). We also see an example of signalling incipient speakership in line 46, which would be (incorrectly) considered a Listener Response if we only looked at the talk immediately surrounding any of the current Listener's contributions. I will discuss both in turn. Donna and PuzzleB are talking about the importance of speaking to other people with Type 1 Diabetes and feeling understood when sharing stories of their daily ups and downs:

(4) Donna and PuzzleB, Diabetes Camp and DAFNE, minutes 16:15–18:00.

31		PuzzleB:	=it's just like 'Oh, being with other people
32			who 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\\.
39		Donna:	[um::-
40		Donna:	and so it was good meeting them-
41		Donna:	(.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
52			there's still the-
53	→	Donna:	yeah: [:-
54		PuzzleB:	[**yeah** and the-
55		PuzzleB:	**yeah** cause that's what happened my daughter
56			after going on the: um camp is that-
			PUZZLEB CONTINUES

PuzzleB has been speaking about her daughter's experience at a diabetes camp and her comment in lines 31/32 refers to this. Her statement is potentially incomplete, signalled prosodically through flat final intonation, and syntactically by 'and' being the last word. However, this is followed by a 0.5 s silence (line 33), and then Donna initiates a new turn, taking over the floor with 'yeah cause I'd never'. As we have seen in extract 2, yeah is often used as a Listener Response. In this position though, it follows a pause and is immediately followed by more talk from Donna rather by talk from PuzzleB, who held the floor earlier, which makes it an instance of incipient speakership rather than a Listener Response (see Jefferson (1984) and Jefferson (1993) on 'yeah' as a preface to immediate on-topic continuations). We see the same with 'oh right' in line 51, and 'yeah' again in lines 54 and 55.

Additionally, we can see one (in this data extremely rare) example of the floor being negotiated on lines 44–54. In line 44 of Donna's talk, potential completion of her turn is projectable. PuzzleB provides a Listener Response in overlap with this utterance (line 45), and then vocalises ‘er:::-’ in the brief break between Donna's talk and the completion of this turn on line 47. PuzzleB immediately follows up her ‘er:::-’ with a question (line 48) and repeats the answer back to Donna in overlap with the final syllable of Donna's turn (lines 49/50). PuzzleB's ‘oh really’ on line 45 is still counted as a Listener Response, while ‘er:::-’ is considered a signal of incipient speakership and claim to the floor, which then leads to a floor change.[3]

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, and eliciting more same-Speaker talk.
2. They *tend* to be placed at TRPs (Ford and Thompson, 1996; Clancy et al., 1996), **but**
3. they do not *have to be* produced at every possible point of occurrence, and their occurrence is not restricted to TRPs.

Now that we have an interactional definition of Listener Responses, we need 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 (Clancy et al., 1996; Ford and Thompson, 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 illustrates that 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. Hence, the best current 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:

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

This rate is calculated for each individual turn, and then multiplied by 100, the mean turn length both in Duncan and Fiske (1977) and in the present study, because the (hypothetical) number of ‘14 responses per 100 interlocutor words’ (i.e. the average length of a turn) is more intuitively interpretable than ‘0.14 responses per interlocutor word’. This builds on Duncan and Fiske (1977), who analysed frequency at the level of the conversation. Calculating Listener Response frequency turn-by-turn allows for a more fine-grained, nuanced analysis. I will return to the theoretical and methodological implications of this approach in more detail in the discussion.

4. Data and quantitative methods

Having already referred to several short examples from the set of recordings that form the basis of this study, I will now briefly introduce the data overall, the coding process, and the quantitative analysis based on this interactional definition of the variable.

4.1. The data

The data consist of 24 dyadic, topic-focused conversations about living with Type 1 Diabetes, each with two participants. There are 10 all-female dyads, 13 mixed dyads, and 1 male-only dyad.[4]

All conversations were coded by me for Listener Responses based on the definition given in the previous section. 10% were also coded by a variationist sociolinguist who had no prior training in interactional approaches like CA. Interrater reliability

[3] We can think of other contexts – for example department meetings or panel discussions in talk shows – in which there is far more and very explicit competition for the floor, often in the form of overlap. Such contributions are not Listener Responses, even though at a purely structural level they stand ‘in’ the ongoing talk and might not lead to an immediate floor change. However, Speakers respond to competitive overlap with turn-holding strategies like for example rush-throughs, rising intonation, or glottal gestures (Ogden, 2001; Local and Kelly, 1986; Levon, 2016). Further, competitive overlap also differs in form; it tends to be louder, faster, and uses different lexical material from Listener Responses (French and Local, 1983; Riou, 2017; Local and Walker, 2012). In contexts where there is more competition for the floor than in the data underlying this study, ‘form’ (as alluded to in point 1 of the summary below) thus becomes important in distinguishing competitive overlap from Listener Responses.

[4] As mentioned earlier, there were 11 female and 5 male participants in 4 groups of 4 participants each. Within each group, each person spoke to the other three participants individually, resulting in the gender-split for dyads outlined above. This imbalance in the gender split of the dyads is owed to the context of the larger project in which the data were collected. This project investigates 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 caution the reader to be particularly careful not to generalise from the one all-male dyad.

(IRR) was very high, with *F-Measure* = 0.97, *Recall* = 0.96, *Precision* = 0.99. This demonstrates how easy it is to apply this interaction-based coding.

Overall, there are 5202 Listener Responses in the dataset, 3848 of which are produced by women, and 1354 by men (Table 1). If we relate this to the gender distribution in the sample, this is not surprising: 33 (or 68%) of the 'Speaker slots'[5] are occupied by women, and only 15 (or 32%) by men. Assuming that Speaking time is evenly distributed and that there is no gender difference in how often Listeners respond, we might expect men to be accountable for roughly a third of all Listener Responses.

Table 1
Cross-tabulation of raw token counts of Listener. Responses by Listener and Speaker gender across all interactions.

	Female Speaker	Male Speaker	Total by gender
Female Listener	2291	1557	3848
Male Listener	1100	254	1354
Total			5202

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 (Fig. 1).

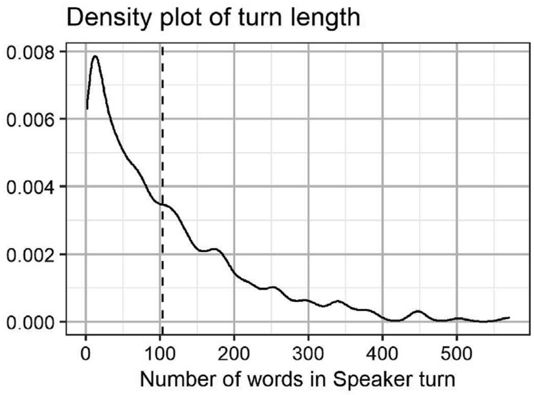

Fig. 1. Density plot showing turn length follows a log-normal distribution, with the mean being at 103 and the median at 75. Turn length ranges from 1 to 570 words.

Fig. 1 shows that turn length ranges from 1 to 570 words[6] (a very extended narrative), with a mean close to 100 words and a median of 75 words. Both turn length and the number of Listener Responses are log-normally distributed. The number of Listener Responses is further skewed towards 0, with a large number of brief turns receiving no Listener Response at all. To reduce skewness and prepare the data for statistical analysis, turn length was centred around the median of 75 words and log-normalised. Listener Response frequency per turn was log normalised with logp1() (see Baayen (2008)). The log1p() function in R calculates log(x + 1) which is necessary for Listener Responses, where the value can be 0.

4.2. Statistical methods

All conversations were coded in ELAN. Data was then exported as a csv and processed and analysed using R and R Studio (R Core Team, 2016; RStudio, 2015).

The second key contribution of this paper is to highlight the potential of Zero-inflated Poisson (ZIP) regression models as a statistical tool which could find much wider application in studies of sociolinguistic variation. Currently, poisson regression is used for example in psycholinguistics (Lo and Andrews, 2015; Rigby et al., 2008), typology (Coupé, 2018), or computationally advanced models of language change (Winter and Wieling, 2016). They are very common in disciplines where the number of

[5] 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-slots' and 48 'Listener-slots'.

[6] Words were separated based on the unnest_tokens() function from the tidytext package (Silge and Robinson, 2017) and counted automatically. This allows for maximal reproducibility of the word counts across all conversations.

events of type x need to be modelled – for example incidence, fertility, or mortality in medicine (Cleophas and Zwinderman (2018), Gagnon et al. (2008), Jackson et al. (2016), and Mouatassim and Ezzahid (2012)).

Zero-Inflated Poisson regression models serve to interpret count data with a high number of zeros which could be related to different predictor variables. Such a 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.

Speaker gender (i.e. the gender of the person doing the talking) is added as a fixed effect. The model also takes into account that each individual behaves differently and that each dyad has its own dynamic (i.e. the individual Listener's behaviour is not fixed) by adding random intercepts for both Listener and dyad.[7]

The same formula was used for both the Zero- and the count model, and the models were fit with the glmmTMB package (Brooks et al., 2017). The model 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 output and a brief comparison to two alternative models with only Listener and only Dyad respectively as random effects are given in the results section.

5. Results

The following sections first show descriptive statistics of observed variation in the number of responses given to any turn between male and female Speakers and Listeners relative to turn length. Then, results from a Zero-inflated Poisson regression model are presented, which suggest that most of the observed variability is due to individual rather than group-level differences.

5.1. Descriptive statistics

Based on plotting observed turn length on the x-axis and number of Listener Responses on the y-axis, Fig. 2 shows that the number of Listener Responses in any given turn increases with turn length, and that there seems to be an effect of both Speaker and Listener gender.[8]

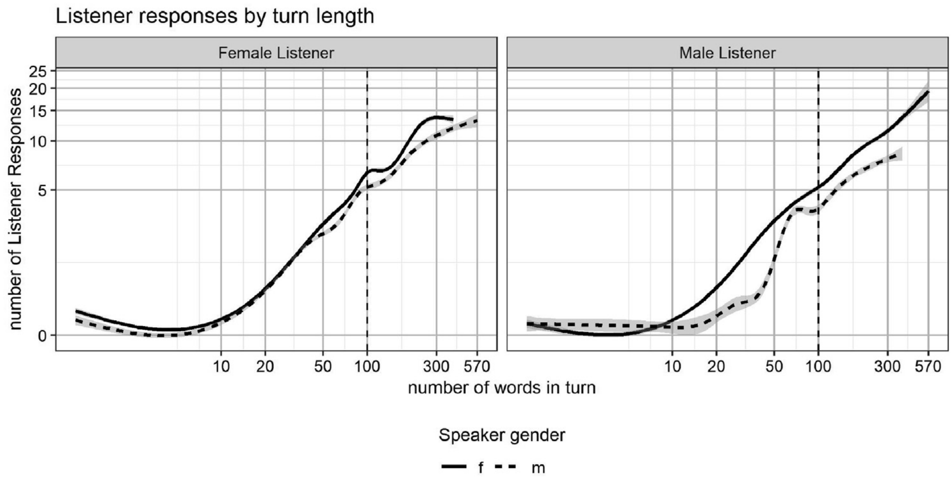

Fig. 2. Number of Listener Responses increases with turn length, but to a different extent depending on Speaker and. Listener gender. Female Listeners produce Listener Responses more frequently than men.

Further, turn length needs to exceed a certain number of words before a Listener Responses is given. Specifically, Listeners respond to turns that are 10 words or longer, with small differences by Listener and Speaker gender. Overall, this plot shows women in the present data produce more Listener Responses than men, especially when listening to a female Speaker.

[7] 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.

[8] The lines were created with ggplot2's geom_smooth() function and simply serve to illustrate the trend. They are not equivalent to the Zero-inflated Poisson regression model presented in the next section.

Further, descriptively, the impact of Speaker gender appears greater for male Listeners in the data at hand. There seems to be cross-gender accommodation in terms of Listener Response frequency.[9]

However, at least some of the observed variation might be driven by individual participants. Fig. 3 shows that there is large within-group variation, especially when the Speaker (turn-holder) is female.

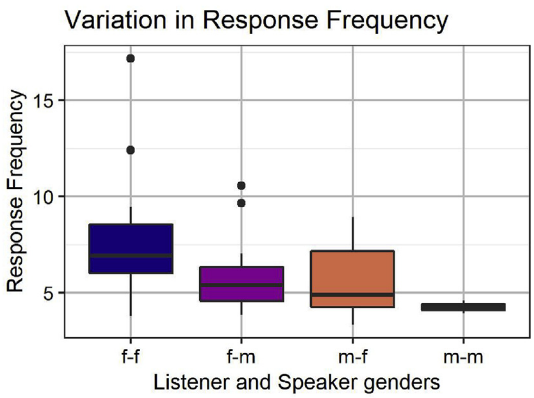

Fig. 3. Boxplot showing variation in the ranges and standard deviations of Listener Response frequency by Listener and Speaker genders.

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. Accordingly, we need to include Listener as a random effect in any statistical model on frequency. To shed light on what is driving this variation, the next section presents a ZIP model that takes into account the random effects of Listener *and* dyad as well as the fixed effects of turn length and Listener and Speaker gender.

5.2. Inferential statistics

A Zero-inflated Poisson regression model was fit to explore the effects of turn length and Listener and Speaker gender, while taking into account random effects due to Dyad and Listener. The model summaries are given in Table 2 for the logit model and Table 3 for the count model. The model output is given in log units initially, but later converted to predicted counts in Tables 4 and 5 for ease of interpretation.

Table 2
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)	−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**

P-values that are statistically significant are highlighted in bold.

Table 3
Fixed Effects of the count model: number of Listener. Responses increases with turn length (strong effect), and male Speakers receive fewer Responses than female ones (marginal effect).

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

P-values that are statistically significant are highlighted in bold.

9 Indeed, in a model of the same structure as the random effects model described in the next section (but without those random effects of Listener and dyad) both Speaker and Listener gender alongside turn length are highly significant ($p < 0.001$), though the effect size varies. The full regression table with brief explanations of the results is given in Appendix A and can be read in parallel to the mixed effects model below.

Table 4
Model prediction for the length of a turn that receives 1. Listener Response: male Listeners wait for longer before they produce the first Response.

	Female Speaker	Male Speaker
Female Listener	11.4	14
Male Listener	12.2	14.8

Table 5
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.

	Female Speaker	Male Speaker
Female Listener	6 (7.8)	5.1 (6.2)
Male Listener	5.5 (5.8)	4.7 (4.5)

Overall, the ZIP model shows that the gender difference is not statistically significant when taking individual and dyad-specific aspects into account. 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 count model Speaker (Variance = 0.016, sd = 0.128) and dyad (Variance = 0.050, sd = 0.223).

Turn length is the most important predictor for how many Listener Responses will be given to a turn conceptually and statistically; it is the only fixed effect which is statistically significant in *both* parts of the model (model summaries are given in Tables 2 and 3). The longer the turn, the less likely it is to receive zero responses, and the number of predicted Listener Responses increases with turn length.

Both parts of the model take ‘female’ as the default gender and therefore report the effect of gender for the male Speakers and Listeners. Listener and Speaker gender alone have no significant effect. Only in its interaction with turn length does Listener gender matter to a small extent, with an effect size of −0.23 ($p = 0.003$): as turn length increases, male Listeners are less likely than female Listeners to produce no Response at all to a turn of the same length.

In the count model, the only clearly significant factor with a strong effect size is turn length (Estimate = 0.57, $p < 0.001$). Speaker gender is marginally significant ($p = 0.048$), but the effect size is far smaller than for turn length (Estimate = −0.16).[10] We can use the model output to calculate how many words Speakers are predicted to utter before the Listener begins to respond (Table 4), bearing in mind that here we are talking about trends rather than statistically significant group-level results.

In the descriptive statistics it looked as though Listeners began responding once the turns exceeded 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. Female Speakers are predicted to produce 11.4 words for a female Listener to do 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.

If we compare the predicted number of Listener Responses to the observed, we get a first intuitive idea of how great the impact of the individual Listener and dyad are. Table 5 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 are given in brackets next to the predicted numbers. The top left cell of Table 5 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.

The comparison between the predicted and observed number of Listener Responses to a 100-word turn demonstrates two things: first, 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 and/or dyads who produce Listener Responses very frequently. Second, it concurs with the observed *trend* 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, though none of these gender differences is statistically significant without including turn length.

The inclusion of Listener *and* dyad as random effects is theoretically motivated: individual participants behave very differently, and each dyad has a unique interactional dynamic that goes beyond a simple addition of Speaker and Listener as

[10] Given the small effect size and marginal significance level, this might well be a Type 1 error.

individual random effects. To assess whether this theoretically motivated model is also statistically the best fit, two alternative models were created: one with Listener only, and one with dyad only as a random effect. The model including both Listener and dyad as random effects is the best fit based on a comparison of AIC values, followed by the model only including dyad. The model only including Listener as a random effect has the lowest fit.[11]

Table B1 (in appendix B) compares which factors are significant in the three models. Turn length is consistently highly significant, while Listener gender is not significant in any of the models. In the model including only Listener as a random effect, all factors but Listener gender are statistically significant, while the model with only dyad as a random effect is very similar to the best fit model including both factors. However, in the count model Speaker gender is significant at $p = 0.021$ compared to only marginally significant at $p = 0.048$ in the best fit model. The interaction between turn length and Listener gender is also significant in the model only including dyad as a random effect ($p = 0.013$), while it is not significant in the final model. This suggests that there is great individual variation in how Listeners respond to male and female Speakers, and to turns of a given length, and this variation is only captured when adding "Listener" as a random effect. Overall, the theoretically motivated inclusion of both Listener *and* dyad as random effects leads to the model with the best fit, and the changes in significance levels from model to model reflect qualitative observations about the data.

To summarise, the descriptive results suggested a strong effect of turn length, and a gender difference, with male Listeners producing and male Speakers receiving fewer responses than the female Listeners and Speakers in this dataset. However, we also noted high within-group variation. The inferential statistics, which account for random effects of individuals and dyads show that turn length has the strongest effect on how many responses are predicted: the longer the turn, the more responses we expect. Further, there is an interaction between Listener gender and turn length for the zero-inflated model ($p = 0.003$), predicting male Listeners are less likely than female Listeners to produce no response as turn length increases, and a marginally significant effect of Speaker gender in the logit model ($p = 0.048$), which predicts male Speakers to receive fewer responses than female Speakers. However, overall the inferential statistics show that most of the variation noted in the descriptive statistics is in fact due to the random effects of individual Listener and dyad.

5.3. *Variation based on frequency operationalisation*

As outlined in the introduction, there are three common approaches to operationalising and normalising frequency: the number of Listener Responses per 100 words in the corpus, the number of Listener Responses in 5 min of talk (neither of them split by Speaker-Listener Roles), the percent of Speaker changes that are Listener Responses (as opposed to floor changes). Several authors have argued that these ways of quantifying are prone to interactional confounds. Given that the quantification proposed in this paper sets out to address this, I briefly compare descriptive statistics for these metrics and the operationalisation I propose.

The boxplots in Fig. 4 represent the gender-composition of any given dyad.[12] The scale of the y-axes varies in the four facets, because the four metrics refer to different types and ranges of numbers.

It is important to remember that the data I worked with is fairly balanced in terms of the distribution of Speaker and Listener roles. Nevertheless, we can see that the first three approaches overstate how often male Listeners respond to female Speakers, and that results differ depending on the frequency operationalisation.

The first facet shows frequency per 100 words in the corpus – conceptually this is equivalent to 'per million words', but using 100 words allows for a direct comparison with the metric I proposed. Because I have more mixed-gender than same-gender dyads, I normalised by each gender-combination rather than the full corpus. The results suggest minimal variation in Listener Response frequency based on Listener gender (f–f, m-f), while male Speakers receive fewer responses from both male and female Listeners (f–m, m–m). This runs counter to Stubbe (1998) and Murphy (2012) who find men to produce more responses per million words in the corpus and do not discuss the potential effect of the interlocutor.

The second facet visualises the number of responses in 5 minutes of interaction. Here, between-group differences are minimal, with no visible difference based on Listener gender (f–f, m-f, m–m), while there seems to be an interlocutor effect for women, who produce more responses in same- than in mixed-gender dyads (f–f vs. f–m). This could be argued to tie in with the previous finding that female Speakers *receive* more responses than male Speakers (Reid, 1995). In my data this effect is only visible for female participants. Oreström (1983) notes no gender difference in frequency per set amount of time, but points out that which tokens are used varies.

The third facet shows the percent of Speaker changes in any given conversation which are Listener Responses rather than floor changes (see section 3.1), used for example in Kogure (2003), who found no gender difference but large individual variability (and in Clancy et al. (1996), but not with respect to gender). Contrary to Kogure (2003), my results suggest an interaction between Speaker and Listener gender: overall, for female Listeners the frequency of responses relative to floor

[11] This is the case based on AIC values in the model output, which are 83830.7 for the model including both random effects, 84952.6 for the model including Dyad only as a random effect, and 87361.5 for the model including Listener only as a random effect. The lowest AIC indicates the best fit. This ranking is reflected in the output of the AIC comparison with AICtab in R, which rates the model with both Listener only as a random effect as the worst (delta AIC = 3530.8), which is improved greatly when including dyad only as a random effect instead (delta AIC = 1121.9), and rated as best when including both factors as random effects (delta AIC = 0).

[12] f–f refers to a woman listening to a woman, f–m to a woman listening to a man, m–m to an all-male dyad, and m-f to a man listening to a woman.

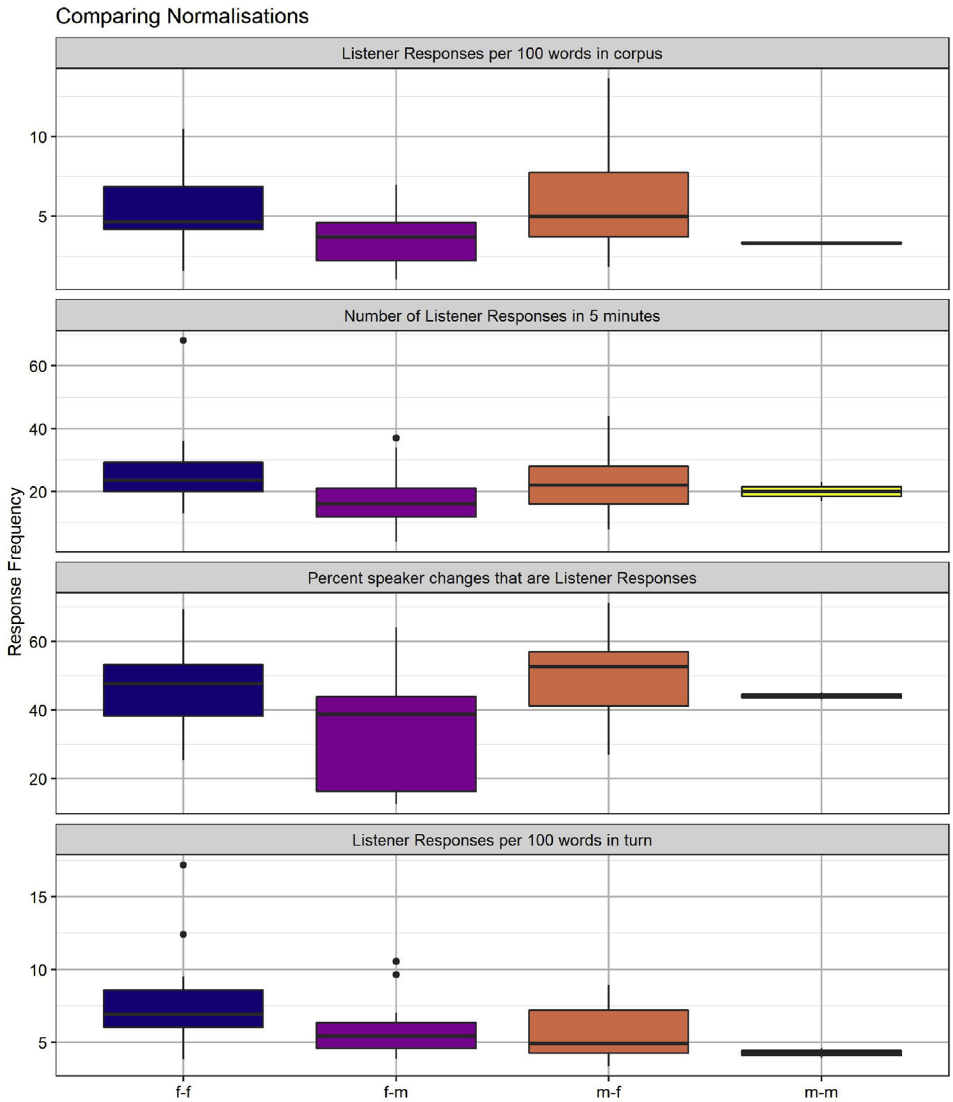

Fig. 4. Facet plot comparing three ways of operationalising. Listener Response frequency to the method outlined in this chapter. 'F–F' refers to a female Listener and a female Speaker, 'F-M' to a female Listener and a male Speaker, and so on. Differences observed vary depending on the method chosen.

changes is lower than for male Listeners (f–f and f–m vs. m-f and m–m). For both male and female Listeners this percentage decreases when listening to a male Speaker (m–m and f–m). The height of the boxes for male and female Listeners in mixed-gender dyads reflects the large individual variability and is a cautionary note to the importance of inferential statistics, which would be difficult to run on this data.

The final facet shows the frequency operationalisation I have proposed. My findings directly contrast with Dixon and Foster (1998) who found male Listeners to produce the highest number of responses towards female Speakers (m-f) in a sample of South African undergraduate students.

This comparison of four different approaches to operationalising frequency shows that the method impacts on the patterns we can observe, and underlines the need for an interactionally accountable approach to quantification.

6. Discussion and conclusion

To summarise, this paper presents a theoretical and methodological approach which allows us to root variation above the level of the phoneme in the local discourse-organisational structure, based on the example of Listener Responses and gender. It defines the variable primarily based on its sequential position and impact, and quantifies it relative to the talk it is responding to, as number of Listener Responses in a turn relative to the length of that turn (in words). It further demonstrates the application of Zero-inflated Poisson regression models as a statistical tool which allows us to model sociolinguistic data with high numbers of zeroes.

This definition of the variable aligns with and builds on the proposal to integrate CA tools in the definition and analysis of discourse-pragmatic variables (Pichler, 2013) and (variational) pragmatic approaches to defining discourse-pragmatic variables based on their positional, derivational, functional, and/or formal equivalence (Barron et al., 2017; Ito and Tagliamonte, 2003; Schneider, 2010; Staley, 2018; Waters, 2016).

Rooting the variable in *interactional* structure adds to this work by offering a new perspective on where the envelope of variation is situated: in the traditional Labovian paradigm, this is within the person producing the variable, and frequency is accordingly quantified relative to that person's linguistic production. This is so taken for granted that whenever I introduce the approach to quantification presented here, I make sure to point out why the number of Listener Responses is related to the *Speaker's* (that is, the *interlocutor's*) turn length rather than the *Listener's* linguistic production. This is not to say that all variables should suddenly be considered as rooted in the interaction, and quantified relative to it, but the approach presented here opens up a new analytical space to treat variables that are, in fact, interactionally structured, in an accountable way.

How we operationalise frequency influences the results we get, and the conclusions we draw. This is evident in section 5.3: the first three approaches confound interactional and social effects, meaning we cannot distinguish whether the gender difference in the *observational* data is due to different ways of Listening or caused by differences in the distribution of Speaker and Listener roles. Furthermore, the high within-group variability highlights the importance of *inferential* statistics to account for random effects. The approach I have proposed addresses these issues: it offers an interactionally accountable way of operationalising frequency, where the interaction is part of the metric rather than a confound, and the Zero-inflated Poisson regression allows for the inclusion of random effects. The ‘random’ variability in the model presented might relate to other social factors, for example culture (Heinz, 2003; O'Keeffe and Adolphs, 2008; White, 1989) or age, as well as individual factors like conversational style (Tannen, 2007) not included in this study. While level of familiarity might also have an impact, it did not vary in the corpus under study; none of the participants had met their interlocutors prior to the recordings.

Different aspects of the methodology presented here could be extended to other variables above the level of the phoneme, for example Neg-tags defined by derivational equivalence (Pichler, 2013) or quotatives (Buchstaller, 2014) defined by functional equivalence. In both contexts, an interactional approach promises a more fine-grained understanding of what the variable and its different variants can do. A CA analysis of different ways of introducing quotation might uncover a relationship between the quotative used, its precise realisation, and the context it occurs in. This could elucidate patterns of variation related to socio-demographic factors. For Neg-tags, a CA analysis of all positions in which Neg-tags *do* occur could help circumscribe the envelope of variation in a new way and open up ways of discovering “absences” of Neg-tags. Here, a deviant-case analysis in which participants orient to the use – or lack of use – of a Neg-tag would be especially interesting as a follow-up to Pichler (2013). The exact aspect of the methodology to be transferred depends, of course, on the variable and the research question (see also recent discussions in Variational Pragmatics on the different levels of variation to be analysed, particularly by Barron (2017) and Staley (2018)).

This paper outlines the first step towards a more interactionally accountable analysis of variables above the level of the phoneme, which integrates CA and variationist approaches. Thus, it naturally opens up several avenues for further work:

Here, we have looked at the overall frequency of Listener Responses relative to the *number of words* in the Speaker's turn. How quickly the Speaker utters these words, and how many pauses and hesitations are included in the turn might also affect the number of responses. This could be explored by including speech rate and rate of articulation in the analysis (see Kendall (2013)) for an excellent discussion of the complexities of operationalising those two variables). Further analyses could also look at the actual position of each Listener Response in the turn (i.e. the first one after 5 words, the next one after 10 more words and so on) and how this changes over the course of the conversation.

The analysis of Listener Responses can be extended to the other levels of pragmatic analysis outlined in Schneider (2010) and Barron et al. (2017) as follows: our interactions are not only structured at the local level under study here (the organisational level), but also with respect to interactional projects like telling a story vs. giving a lecture (the interactional level). This has been argued to impact discourse-level variation (see for example Buchstaller (2011) and Guardiola et al. (2012)), and developing a theoretical and methodological approach that encompasses both levels will be another important contribution to make.

Furthermore, it is evident in the examples presented earlier that Listeners can do a number of different actions, and responses can take varying forms (the actional and the formal level). Both Fellegy (1995) and Oreström (1983) have found no gender difference in frequency, but variation in the *types of responses* or their *form*. Thus, an interactionally rooted coding scheme for the different actions Listeners can do is needed, followed by a quantitative analysis taking into account variation in the action types, and an analysis of *how* Listeners respond in terms of the linguistic material they use in their responses. This is developed in Eiswirth (2020b).

Future work should also look into the role of situational context – as noted in section 3.2, the data under study here stem from very attentive, involved Listeners that could be argued to be ‘doing being a good Listener’. We might find very different patterns of Listener Responses in everyday-interactions or in institutional talk.

Thus, this paper presents an innovative methodology for rooting the definition, quantification, and analysis of a variable above the level of the phoneme – Listener Responses – in interactional structure, and provides a starting point for further work towards a closer integration of interactional and variationist analyses to also include formal, functional, broader interactional, and situational aspects.

Declaration of competing interest

None.

Acknowlegements

I would like to thank my PhD supervisors Dr Lauren Hall-Lew, Dr Josef Fruehwald, and Dr Joseph Gafaranga, and my examiners Professor Dr Isabelle Buchstaller and Dr Sumin Zhao. I would also like to thank conference audiences at DiPVaC4 in Helsinki and ICLaVE10 in Leeuwarden, and colleagues at the Universities of Edinburgh and Glasgow for their feedback on earlier versions of this analysis. Finally, thank you to the anonymous reviewers of this paper for their detailed and generous comments that have helped shape this final version. All remaining errors are my own. PhD fellowship by the German National Academic Foundation,PhD fellowship by the Arts and Humanities Research Council,PhD fellowship by the German Academic Exchange Service.

Appendix A

Model output of the *fixed effects* ZIP model using glmmTMB and the same formula as the mixed effects model presented in the analysis: number of Listener Responses as predicted by the factors turn length, Listener gender, Speaker gender, and the interaction between turn length and Listener gender.

Table A1
Fixed Effects of the logit model: the longer the turn, the more likely there is to be a Listener Response. Male Listeners let more words pass before doing the first Response, and male Speakers utter more words before their Listener produces the first Response. With increasing turn length, male Listeners are less likely than female Listeners to produce no Response at all.

Fixed Effect	Estimate	Std. Error	Z-Value	p-value
(Intercept)	−3.79	0.094	−39.28	<0.001
words in turn (centred)	−0.85	0.026	−32.52	<**0.001**
Listener (male)	0.27	0.123	2.17	**0.003**
Speaker (male)	0.73	0.101	7.22	<**0.001**
words in turn (centred) x Listener (male)	−0.35	0.051	−6.82	<**0.001**

P-values that are statistically significant are highlighted in bold.

Table A2
Fixed Effects of the count model: number of Listener Responses increases with turn length. Overall, male Listeners produce and male Speakers receive fewer Responses than female Listeners or Speakers. Male Listeners produce fewer responses relative to the same increase in Speaker's turn length than female Listeners.

Fixed Effect	Estimate	Std. Error	Z-Value	p-value
(Intercept)	1.61	0.006	263.39	<0.001
words in turn (centred)	0.52	0.004	125.79	<**0.001**
Listener (male)	−0.06	0.010	−5.87	<**0.001**
Speaker (male)	−0.19	0.006	−27.73	<**0.001**
words in turn (centred) x Listener (male)	−0.03	0.07	−4.24	<**0.001**

P-values that are statistically significant are highlighted in bold.

Appendix B

Table B1
Schematic comparison of models including different random effects: the model including both Listener and dyad as random effects is the model with the best fit.

	Listener only as random effect	Dyad only as random effect	Both Listener and dyad as random effects
	Count Model		
words in turn (centred)	$p < 0.001$	$p < 0.001$	$p < 0.001$
Listener (male)	(ns)	(ns)	(ns)
Speaker (male)	$p < 0.001$	$p = 0.021$	$p = 0.048$

Table B1 (*continued*)

	Listener only as random effect	Dyad only as random effect	Both Listener and dyad as random effects
words in turn (centred) x Listener (male)	$p = 0.015$	$p = 0.013$	(ns)
	Logit Model		
words in turn (centred)	$p < 0.001$	$p < 0.001$	$p < 0.001$
Listener (male)	(ns)	(ns)	(ns)
Speaker (male)	$p < 0.001$	(ns)	(ns)
words in turn (centred) x Listener (male)	$p < 0.001$	(ns)	$p = 0.003$

References

Baayen, R.H., 2008. Analyzing Linguistic Data: A Practical Introduction to Statistics Using R. Cambridge University Press. https://doi.org/10.1558/sols.v2i3.471.

Barron, A., 2017. Variational pragmatics. In: The Routledge Handbook of Pragmatics, pp. 91–104. https://doi.org/10.4324/9781315668925-10.

Barron, A., Gu, Y., Steen, G. (Eds.), 2017. The Routledge Handbookof Pragmatics, first ed. Routledge, London. https://doi.org/10.4324/9781315668925.

Barron, A., Schneider, K.P., 2009. Variational pragmatics: studying the impact of social factors on language use in interaction. Intercult. Pragmat. 6, 425–442. https://doi-org.ezproxy.is.ed.ac.uk/10.1515/IPRG.2009.023.

Bavelas, J.B., Coates, L., Johnson, T., 2000. Listeners as co-narrators. J. Pers. Soc. Psychol. 79, 941–952. https://doi.org/10.1037/0022-3514.79.6.941.

Brunner, L.J., 1979. Smiles can be back channels. J. Pers. Soc. Psychol. 37, 728–734. https://doi.org/10.1037/0022-3514.37.5.728.

Brooks, M.E., Kristensen, K., van Benthem, K.J., Magnusson, A., Berg, C.W., Nielsen, A., Skaug, H.J., Mächler, M., Bolker, B.M., 2017. glmmTMB balances speed and flexibility among packages for zero-inflated generalized linear mixed modeling. R J. 9, 378–400. https://doi.org/10.3929/ETHZ-B-000240890.

Buchstaller, I., 2006. Diagnostics of age-graded linguistic behaviour: the case of the quotative system. J. Sociol. 10, 3–30. https://doi.org/10.1111/j.1360-6441.2006.00315.x.

Buchstaller, I., 2008. The localization of global linguistic variants. English World-Wide 29, 15–44. https://doi.org/10.1075/eww.29.1.03buc.

Buchstaller, I., 2011. Quotations across the generations: a multivariate analysis of speech and thought introducers across 5 decades of Tyneside speech. Corpus Linguist. Linguistic Theory 7, 59–92. https://doi.org/10.1515/CLLT.2011.004.

Buchstaller, I., 2014. Quotatives: New Trends and Sociolinguistic Implications. John Wiley & Sons, Inc, Oxford, UK. https://doi.org/10.1002/9781118584415.

Clancy, P.M., Thompson, S.A., Suzuki, R., Tao, H., 1996. The conversational use of reactive tokens in English, Japanese, and Mandarin. J. Pragmat. 26, 355–387. https://doi.org/10.1016/0378-2166(95)00036-4.

Cleophas, T.J., Zwinderman, A.H., 2018. Modern Bayesian Statistics in Clinical Research, Modern Bayesian Statistics in Clinical Research. Springer International Publishing, Cham. https://doi.org/10.1007/978-3-319-92747-3.

Clift, R., 2016. Conversation Analysis. Cambridge University Press. https://doi.org/10.1017/9781139022767.

Coupé, C., 2018. Modeling linguistic variables with regression models: addressing non-Gaussian distributions, non-independent observations, and non-linear predictors with random effects and generalized additive models for location, scale, and shape. Front. Psychol. 9, 513. https://doi.org/10.3389/fpsyg.2018.00513.

Dines, E.R., 1980. Variation in discourse: and stuff like that. Lang. Soc. 9, 13–31. https://doi.org/10.1017/S0047404500007764.

Dixon, J.A., Foster, D.H., 1998. Gender, social context, and backchannel responses. J. Soc. Psychol. 138, 134–136. https://doi.org/10.1080/00224549809600364.

Duncan, S., Fiske, D.W., 1977. Face-to-face Interaction: Research, Methods, and Theory. L. Erlbaum Associates.

Eiswirth, M., 2020a. It's all About the Interaction – Listener Responses as a Discourse-Organisational Variable. University of Edinburgh. https://doi.org/10.7488/era/64.

Eiswirth, M., 2020b. Modeling Listener Responses. In: Peterson, E., Hiltunen, T., Kern, J. (Eds.), Dynamics and Innovations in Discourse-Pragmatic Variation and Change. Cambridge University Press, Cambridge.

French, P., Local, J., 1983. Turn-competitive incomings. J. Pragmat. 7, 17–38. https://doi.org/10.1016/0378-2166(83)90147-9.

Fellegy, A.M., 1995. Patterns and functions of minimal response. Am. Speech 70, 186–199. https://doi.org/10.2307/455815.

Ford, C.E., Thompson, S.A., 1996. Interactional units in conversation: syntactic, intonational, and pragmatic resources for the management of turns. In: Ochs, E., Schegloff, E.A., Thompson, S.A. (Eds.), Interaction and Grammar. Cambridge University Press, Cambridge, pp. 134–184. https://doi.org/10.1017/CBO9780511620874.003.

Gagnon, D.R., Doron-LaMarca, S., Bell, M., O'Farrell, T.J., Taft, C.T., 2008. Poisson regression for modeling count and frequency outcomes in trauma research. In: Journal of Traumatic Stress. John Wiley & Sons, Ltd, pp. 448–454. https://doi.org/10.1002/jts.20359.

Gardner, R., 1998. Between speaking and listening: the vocalisation of understandings. Appl. Linguist. 19, 204–224. https://doi.org/10.1093/applin/19.2.204.

Guardiola, M., Bertrand, R., Espesser, R., Rauzy, S., 2012. Listener's responses during storytelling in French conversation. INTERSPEECH.

Heinz, B., 2003. Backchannel responses as strategic responses in bilingual speakers' conversations. J. Pragmat. 35, 1113–1142. https://doi.org/10.1016/S0378-2166(02)00190-X.

Heritage, J., 1984. A change-of-state token and aspects of its sequential placement. In: Atkinson, J.M., Heritage, J. (Eds.), Structures of Social Action: Studies in Conversation Analysis. Cambridge University Press, Cambridge, pp. 299–345. https://doi.org/10.1017/CBO9780511665868.

Heritage, J., 1999. Conversation analysis at century's end: practices of talk-in-interaction, their distributions, and their outcomes. Res. Lang. Soc. Interact. 32, 69–76. https://doi.org/10.1207/S15327973RLSI321&2_9.

Ito, R., Tagliamonte, S., 2003. Well weird, right dodgy, very strange, really cool: layering and recycling in English intensifiers. Lang. Soc. 32, 257–279. https://doi.org/10.1017/S0047404503322055.

Jackson, C., Mostowy, J.H., Stagg, H.R., Abubakar, I., Andrews, N., Yates, T.A., 2016. Working conditions and tuberculosis mortality in England and Wales, 1890-1912: a retrospective analysis of routinely collected data. BMC Infect. Dis. 16, 215. https://doi.org/10.1186/s12879-016-1509-z.

Jefferson, G., 1993. Caveat speaker: preliminary notes on recipient topic-shift implicature. Res. Lang. Soc. Interact. 26, 1–30. https://doi.org/10.1207/s15327973rlsi2601_1.

Jefferson, G., 1984. Notes on a systematic deployment of the acknowledgement tokens "yeah"; and "mm hm.". Pap. Linguist 17, 197–216. https://doi.org/10.1080/08351818409389201.

Kendall, T., 2013. Speech Rate, Pause and Sociolinguistic Variation. https://doi.org/10.1057/9781137291448.

Kogure, M., 2003. Gender Differences in the Use of Backchannels: Do Japanese Men and Women Accommodate to Each Other? University of Arizona. ProQuest Dissertations Publishing.

Labov, W., 1972. Sociolinguistic Patterns. University of Pennsylvania Press.

Labov, W., 1963. The social motivation of a sound change. Word 19, 273–309. https://doi.org/10.1080/00437956.1963.11659799.

Lavandera, B.R., 1978. Where does the the sociolinguistic variable stop. Lang. Soc. 7, 171–182. https://doi.org/10.2307/4166996.

Levon, E., 2016. Gender, interaction and intonational variation: the discourse functions of high rising terminals in London. J. Socioling. 20, 133–163. https://doi.org/10.1111/josl.12182.
Local, J., Kelly, J., 1986. Projection and Silences: notes on a phonetic and conversational structure. Hum. Stud. 9, 185–204.
Local, J., Walker, G., 2012. How phonetic features project more talk. J. Int. Phon. Assoc. 42, 255–280. https://doi.org/10.1017/s0025100312000187.
Lo, S., Andrews, S., 2015. To transform or not to transform: using generalized linear mixed models to analyse reaction time data. Front. Psychol. 6, 1171. https://doi.org/10.3389/fpsyg.2015.01171.
Mandelbaum, J., 2012. Storytelling in conversation. In: The Handbook of Conversation Analysis. John Wiley & Sons, Ltd, Chichester, UK, pp. 492–507. https://doi.org/10.1002/9781118325001.ch24.
Maynard, S.K., 1997. Analyzing interactional management in native/non-native English conversation: a case of listener response. Int. Rev. Appl. Ling. Lang. Teach. 35, 24.
Maynard, S.K., 1990. Conversation management in contrast: listener response in Japanese and American English. J. Pragmat. 14, 397–412. https://doi.org/10.1016/0378-2166(90)90097-W.
McCarthy, M., 2003. Talking back: ' 'Small" interactional response tokens in everyday conversation. Res. Lang. Soc. Interact. 36, 33–63. https://doi.org/10.1207/S15327973RLSI3601_3.
Mouatassim, Y., Ezzahid, E.H., 2012. Poisson regression and Zero-inflated Poisson regression: application to private health insurance data. Eur. Actuar. J. 2, 187–204. https://doi.org/10.1007/s13385-012-0056-2.
Murphy, B., 2012. Exploring response tokens in Irish English — a multidisciplinary approach: integrating variational pragmatics, sociolinguistics and corpus linguistics. Int. J. Corpus Linguist. 17, 325–348. https://doi.org/10.1075/ijcl.17.3.02mur.
O'Keeffe, A., Adolphs, S., 2008. Response tokens in British and Irish discourse: corpus, context and variational pragmatics. In: Schneider, K.P., Barron, A. (Eds.), Variational Pragmatics. John Benjamins, Amsterdam, pp. 69–98.
Oreström, B., 1983. Turn-taking in English Conversation. Gleerup, Lund.
Ogden, R., 2001. Turn transition, creak and glottal stop in Finnish talk-in-interaction. J. Int. Phon. Assoc. 31, 139–152. https://doi.org/10.1017/S0025100301001116.
Pichler, H., 2013. The Structure of Discourse-Pragmatic Variation, Studies in Language Variation. John Benjamins Publishing Company, Amsterdam. https://doi.org/10.1075/silv.13.
Pichler, H., 2010. Methods in discourse variation analysis: reflections on the way forward. J. Sociol. 14, 581–608. https://doi.org/10.1111/j.1467-9841.2010.00455.x.
R Core Team, 2016. R: A Language and Environment for Statistical Computing.
Reid, J., 1995. A study of gender differences in minimal responses. J. Pragmat. 24, 489–512. https://doi.org/10.1016/0378-2166(94)00066-N.
Riou, M., 2017. The Prosody of topic transition in interaction: pitch register variations. Lang. Speech. https://doi.org/10.1177/0023830917696337.
Rigby, R.A., Stasinopoulos, D.M., Akantziliotou, C., 2008. A framework for modelling overdispersed count data, including the Poisson-shifted generalized inverse Gaussian distribution. Comput. Stat. Data Anal. 53, 381–393. https://doi.org/10.1016/j.csda.2008.07.043.
RStudio, T., 2015. RStudio. Integrated Development Environment for R.
Sacks, H., Schegloff, E.A., Jefferson, G., 1974. A simplest systematics for the organization of turn-taking for conversation. Language 50, 696–735 (Baltim).
Silge, J., Robinson, D., 2017. Text Mining with R, October. https://doi.org/10.1007/978-1-4614-4151-9.
Sankoff, G., 1973. Above and beyond phonology in variable rules. In: Bailey, C.J.N., Shuy, R.W. (Eds.), New Ways of Analyzing Variation in English. Georgetown University Press, Washington, pp. 107–118.
Schegloff, E.A., 1993. Reflections on quantification in the study of conversation. Res. Lang. Soc. Interact. 26, 99–128. https://doi.org/10.1207/s15327973rlsi2601_5.
Schegloff, E.A., 1982. Discourse as an interactional achievement: some uses of 'uh huh'and other things that come between sentences. Anal. discourse Text Talk 71–93.
Schneider, K.P., 2010. Variational pragmatics. In: Fried, M., Östman, J.-O., Verschueren, J. (Eds.), Variation and Change: Pragmatic Perspectives. John Benjamins Publishing Company, pp. 91–104. https://doi.org/10.4324/9781315668925.
Schneider, K.P., Barron, A., 2008. Variational Pragmatics: A Focus on Regional Varieties in Pluricentric Languages. John Benjamins, Amsterdam. https://doi.org/10.1075/pbns.178.02sch.
Schweitzer, A., Lewandowski, N., 2012. Accommodation of Backchannels in Spontaneous Speech, the Booklet of the International Symposium on Imitation and Convergence in Speech.
Sidnell, J., 2012. Basic conversation analytic methods. In: Stivers, T., Sidnell, J. (Eds.), The Handbook of Conversation Analysis. John Wiley & Sons, Ltd, Chichester, UK, pp. 77–99. https://doi.org/10.1002/9781118325001.ch5.
Staley, L., 2018. Socioeconomic Pragmatic Variation. John Benjamins Publishing Company.
Stubbe, M., 1998. Are you listening? Cultural influences on the use of supportive verbal feedback in conversation. J. Pragmat. 29, 257–289. https://doi.org/10.1016/S0378-2166(97)00042-8.
Tannen, D., 2007. Talking Voices: Repetition, Dialogue and Imagery in Conversational Discourse, Studies in Interactional Sociolinguistics. Cambridge University Press, Cambridge.
Terkourafi, M., 2011. The pragmatic variable: toward a procedural interpretation. Lang. Soc. 40, 343–372. https://doi.org/10.1017/S0047404511000212.
Tottie, G., 1991. Conversational Style in British and American English: the case of backchannels. In: Aijmer, K., Altenberg, B. (Eds.), English Corpus Linguistics1. Longman, London and New York, pp. 254–271.
Waters, C., 2016. Practical strategies for elucidating discourse-pragmatic variation. In: Pichler, H. (Ed.), Discourse-Pragmatic Variation and Change in English: New Methods and Insights. Cambridge University Press, Cambridge, pp. 41–56. https://doi.org/10.1017/CBO9781107295476.003.
White, S., 1989. Backchannels across cultures: a study of Americans and Japanese. Lang. Soc. 18, 59–76.
Wilkinson, S., Kitzinger, C., 2014. Conversation analysis in language and gender studies. In: The Handbook of Language, Gender, and Sexuality, second ed. John Wiley & Sons, Inc, Hoboken, US, pp. 141–160. https://doi.org/10.1002/9781118584248.ch7.
Winter, B., Wieling, M., 2016. How to analyze linguistic change using mixed models, growth curve analysis and generalized additive modeling. J. Lang. Evol. 1, 7–18. https://doi.org/10.1093/jole/lzv003.
Wong, D., Kruger, H., 2018. That's right A multifactorial analysis of the selection of backchannel structures in British English. In: Brezina, V., Love, R., Aijmer, K. (Eds.), Corpus Approaches to Contemporary British Speech: Sociolinguistic Studies of the Spoken BNC2014. Routledge, p. 272.
Yngve, V.H., 1970. On getting a word in edgewise. In: Chicago Linguistics Society, 6th Meeting, pp. 567–578.

Mirjam Eiswirth's work focuses on integrating interactional and quantitative approaches to defining the sociolinguistic variable and analysing sociolinguistic variation. She works as a postdoctoral researcher at the Sociolinguistics Lab at the University Duisburg-Essen, and has a PhD in Linguistics and English Language from the University of Edinburgh.