Feedback Relevance Spaces: Interactional Constraints on Processing Contexts in Dynamic Syntax

Christine Howes[1] [ID] · **Arash Eshghi[2]**

Abstract

Feedback such as backchannels and clarification requests often occurs subsententially, demonstrating the incremental nature of grounding in dialogue. However, although such feedback *can* occur at any point within an utterance, it typically does not do so, tending to occur at Feedback Relevance Spaces (FRSs). We present a corpus study of acknowledgements and clarification requests in British English, and describe how our low-level, semantic processing model in Dynamic Syntax accounts for this feedback. The model trivially accounts for the 85% of cases where feedback occurs at FRSs, but we also describe how it can be integrated or interpreted at non-FRSs using the predictive, incremental and interactive nature of the formalism. This model shows how feedback serves to continually realign processing contexts and thus manage the characteristic divergence and convergence that is key to moving dialogue forward.

Keywords Backchannels · Clarification requests · Dynamic syntax · Interaction

1 Introduction

As shown in Example 1,[1] even in broadly monological contexts (e.g. lectures, storytelling; Rühlemann 2007), dialogue is co-constructed by multiple interlocutors. Listeners provide frequent feedback to demonstrate whether or not they have *grounded* the conversation thus far (Clark 1996), i.e. whether something said can be taken to be understood. To achieve this grounding, we produce relevant next turns, or *backchannels*

---

[1] Examples are all taken from the British National Corpus (BNC: Burnard 2000).

✉ Christine Howes
  christine.howes@gu.se

  Arash Eshghi
  a.eshghi@hw.ac.uk

[1] Department of Philosophy, Linguistics and Theory of Science, University of Gothenburg, Gothenburg, Sweden

[2] School of Mathematical and Computer Sciences, Heriot-Watt University, Edinburgh, UK

(e.g. 'mm', as in Example 1:211, or 'yeah' Example 1:194) including non-linguistic cues (e.g. nods).[2] Other responses, such as *clarification requests*[3] (e.g. Example 1:203) indicate processing difficulties or lack of coordination and signal a need for repair (Purver 2004; Bavelas et al. 2012).

Example 1 BNC file KB2:191–213[4]

J	191	Oh we've had er, there was this wait Chris rung us up, she's coming over in April
A	192	Oh is she?
J	193	She's coming over before David
A	194	Yeah
J	195	Have a few weeks [well]
A	196	[Yeah]
J	197	She's having [seven weeks]
A	198	[Oh yeah is she?]
J	199	Ain't she?
	200	And then David's coming over and having three weeks, so she'll be over seven weeks
A	201	Oh that'll be lovely that
J	202	So I said to him
A	203	with the kids?
J	204	Yeah, I said yeah, I mean it is, I said, you come tomorrow if you like bedrooms are ready, but you see she'll be staying at her mum's some time
A	205	Yeah
J	206	She'll be staying here and at her mum's [but er]
A	207	[Yeah]
J	208	[her mum or er]
A	209	[Ooh that'll be nice]
J	210	Her mum really she's got a lot on, she'll have a lot on cos she's got to prepare for that wedding, you know what you're like when you, [you've got]
A	211	[Mm]
J	212	You know if you want, want to be doing things
		[don't you get out of house and that]
A	213	[Yeah, pre preparing for a wedding, yeah] ⟨pause⟩ aye

[2] Although we believe that our analysis should also apply to non-verbal feedback, in this paper, due to the nature of the materials available in the BNC we focus on verbal feedback. As pointed out to us by Mark Dingemanse, one of our reviewers, this may not be straightforward due to differences in the temporal properties of non-verbal feedback, however this does not necessarily mean that continuous feedback (such as an increasingly puzzled expression) does not relate to the utterance in progress at an FRS. See also Lawler et al. (2017) for a related treatment of gestures which are 'asynchronous' to the speech with which they co-refer.

[3] These are essentially the same as what have been called next turn repair initiators (NTRIs) in the Conversation Analysis literature (Schegloff et al. 1977, and others following), but here we follow Purver (2004) in calling them clarification requests throughout in keeping with the more computationally oriented spirit of this paper.

[4] Overlapping talk is shown in aligned square brackets.

As they do not convey any propositional truth conditional content, backchannels are not usually accounted for by traditional linguistic theories, in which they are considered semantically 'empty'. They are thus generally considered part of the 'performance' as opposed to one's (core) linguistic 'competence' (Chomsky 1965), and therefore outside the remit of the grammar.

Similarly, clarification requests (CRs)—along with a swathe of other highly contextual dialogue phenomena—have been systematically left aside as outside the scope of core grammar, and until relatively recently ignored as performance phenomena (see Purver 2004; Fernández 2006; Gargett et al. 2009; Ginzburg 2012; Kempson et al. 2015, 2016, for attempts to integrate them).

These trends notwithstanding, an incremental, dynamic, action-based grammar framework such as Dynamic Syntax (DS: Kempson et al. 2001; Cann et al. 2005; Kempson et al. 2016) allows us to to provide a unitary account of such so-called "performance errors" within the grammar itself, which highlights the primacy of language use in interaction as opposed to an abstract linguistic code (Kempson et al. 2016). This shift of emphasis allows DS to offer parsimonious explanations of otherwise puzzling syntactic phenomena (e.g. clitics; Bouzouita 2008; Chatzikyriakidis and Kempson 2011), as well as accounting for interactive dialogue phenomena (which do not typically occur in written language) such as self-repair (Hough and Purver 2012; Hough 2015) and split utterances (Purver et al. 2010; Kempson et al. 2016; Howes 2012).

In this paper, we present a corpus study of backchannels and clarification requests in British English dialogue[5] that shows that 85% of this feedback occurs at places that the DS model of dialogue predicts they ought to be licensed—so-called *Feedback Relevance Spaces* (FRS; Howes and Eshghi 2017). We further examine those cases that do not apparently occur at an FRS, dividing these into late and early cases. We then describe how, due to the interactive, predictive and incremental nature of the DS formalism, the model can account for this feedback.

2 Background

2.1 Repair Avoidance

Backchannels are usually considered to be "positive" feedback, signalling understanding, with clarification requests taken as "negative" feedback indicating a problem of misunderstanding or misalignment. However, these descriptions mask some parallelisms between different types of feedback, and may miss important insights about the role of feedback.

According to (Schegloff 1982, p. 88), the parallelism between backchannels and clarification requests (and where they may occur, see Sect. 2.2) is not incidental: "…'uh huh', nods, and the like, in passing the opportunity to do a full turn at talk, can be seen to be passing an opportunity to initiate repair on the immediately preceding talk".

---

[5] We hope that our claims in this paper generalise across languages and language families, but this is a question for future work.

This inversion of the usual assumptions around how understanding is reached in dialogue, such that the possibility of misunderstanding is ever present (rather than assuming that the default case is perfectly understood conversational exchanges) prioritises the management of potential misunderstanding in dialogue, with backchannels acquiring their myriad functions as a direct consequence, depending on the action in progress when the backchannel is produced. For example, if the speaker is telling a story and the hearer indicates no need for repair, the backchannel may function as a *continuer*; if giving directions, a backchannel may *acknowledge* identification of a landmark; and if offering an opinion, a backchannel may indicate *agreement*.

"Note that, if tokens such as 'uh huh' operate to pass an opportunity to initiate repair, the basis seems clear for the ordinary inference that the talk into which they are interpolated is being understood, and for the treatment in the literature that they signal understanding. It is not that there is a direct semantic convention in which 'uh huh' equals a claim or signal of understanding. It is rather that devices are available for the repair of problems of understanding the prior talk, and the passing up of those opportunities, which 'uh huh' can do, is taken as betokening the absence of such problems" (Schegloff 1982).

This position—supported by experimental evidence (Mills 2007; Mills and Healey 2006; Healey et al. 2018; Howes et al. 2012b)—means that rather than treating backchannels both as multiply ambiguous (as reported in the literature—see Fujimoto (2007) for a review), and as completely the opposite to clarification requests, we can hope to unify them. Specifically, what unifies these interaction devices as *feedback* is that they are all *procedural* mechanisms for managing the divergence and convergence that is characteristic of dialogue.

On this view, clarification requests are the canonical signal of the divergence of individual takes on the dialogue (characterised as misunderstanding), with backchannels a weak signal indicating no obvious problems, and assumed convergence. Different backchannels, especially those which express the listener's stance may be taken to be stronger evidence for convergence, in line with findings that listeners' choice of generic (e.g. 'mm') or specific ('crikey!') backchannels shapes the speaker's narrative (Bavelas et al. 2000; Tolins and Fox Tree 2014). Relevant next turns and compound contributions (dialogue contributions that continue or complete an earlier contribution which includes so-called *joint* or *split utterances*; Howes 2012) offer an especially strong signal of convergence with even 'subversive' or 'hostile' continuations (Bolden et al. 2019; Gregoromichelaki et al. 2011) indicating convergence in parsing terms, while exploiting the mechanisms offered by the grammar to signal divergence in meaning.

In Dynamic Syntax terms, as we shall see below (Sect. 3.4), these procedural devices all serve to identify divergence and manage the subsequent realignment of semantic processing pathways. It is in such unified, procedural terms that Eshghi et al. (2015) characterise clarification requests and backchannels, with Howes and Eshghi (2017) extending this model to capture their positional distribution or placement. In the present paper, we focus on an empirical investigation of the placement of *backchannels* and *clarification requests*; and for the reasons outlined above (and those discussed in Fujimoto 2007), we will refer to these collectively as feedback unless we want to emphasise their differences.

2.2 Feedback Relevance Spaces

As seen in Example 1:196, backchannels do not just occur at the ends of sentences or turns, but can be produced before a complete sentence has been uttered by one's interlocutor, and the same can be true for other feedback such as clarification requests (as in Example 2). Tracking understanding, via grounding or repair initiation, thus occurs incrementally, before a complete proposition has been produced or processed.

Example 2 BNC file G4K:84–86

J	84	If you press N
S	85	N?
J	86	N for name, it'll let you type in the docu- document name.

However, despite evidence that speaker switch *can* occur at any point in a turn, even within syntactic constituents (Purver et al. 2009; Howes et al. 2011), feedback does not appear to be appropriate just anywhere. Studies using different paradigms such as avatar studies (Poppe et al. 2011) or audio of dialogues with backchannels moved from their actual position (Kawahara et al. 2016) suggests that randomly placed backchannels disrupt the flow of dialogue, are rated as less natural and decrease rapport. Even young children are alert to this—they assess a robot which produces backchannels randomly as a less attentive listener (Park et al. 2017). There is also experimental evidence that mid-sentence CRs (as in Example 2) are more disruptive if they are produced mid-constituent rather than between constituents—leading to more frequent restarts of the interrupted sentence (Eshghi et al. 2010; Healey et al. 2011). This evidence strongly suggests that there are places within and between turns where feedback is relevant.

These *Feedback Relevance Spaces* (FRSs; Howes and Eshghi 2017) are analogous to, but more common than, transition relevance places (TRPs)—places where the turn may shift between speakers (Sacks et al. 1974) and are an extension of *Backchannel Relevance Spaces* (Heldner et al. 2013). Heldner et al. (2013) annotated and compared actual backchannels in Swedish spontaneous dialogue with potential backchannels using subjects acting as third-party listeners. They found lots of individual variation in where subjects chose to place backchannels, and "…on average 3.5 times more backchannel relevance spaces than actual backchannels" indicating that feedback is optional at these points. However, there is no information on what unites these Backchannel Relevant Spaces (e.g. in terms of parts of speech etc—see Sect. 2.3 below), which is necessary if our goal is to *explain* them (Eshghi et al. 2015; Howes and Eshghi 2017).

We should emphasise that the concept of FRSs is, like TRPs, intended as modelling *constraints* on the distributions of the phenomena in question—respectively feedback and speaker transitions—rather than determining them fully. The actual distributions of backchannels and CRs depend on many factors including dialogue genre, topic and task each demanding different levels of mutual understanding from the participants and even the specific feedback tokens appropriate at each FRS. For clarification requests, open-class CRs (Drew 1991) such as 'huh?' or 'what?' are likely to occur after a turn

is complete, whilst other restricted forms of repair initiation, such as reprise fragments ('Bill?') or sluices ('who?') (Purver 2004) can occur mid-turn, e.g. after a noun phrase, as in Example 2. Similarly, for backchannels, 'mhm' is more appropriate than 'okay' before the speaker's turn is complete (Bangerter and Clark 2003). We do not make such fine-grained distinctions in this paper.

Thus, the claim that CRs and backchannels occur at FRSs does not imply that they are the same or that they will follow the same distributional patterns (in fact they do not as we shall see in Sect. 5). The claim is rather that they are subject to the same processing mechanisms and thus do not flout the constraints inherent therein, which, as we argue below, originate in the grammar[6]—see Sect. 3.

2.3 Corpus Studies

As we are interested in the precise placement of feedback (and whether the model presented in Howes and Eshghi (2017) and outlined in Sect. 3.4 can account for it) there are a number of relevant corpus studies, which we outline below, before presenting our own from the British National Corpus (BNC) and how it compares to the studies discussed here.

One of the earliest quantificational studies of feedback is that described in Duncan (1972, 1974). Although based on a limited corpus of two dyadic dialogues (with one person appearing in both dialogues), this study presents a detailed multimodal annotation of backchannel responses. Of 71 total backchannels, 69 contained both vocal and visual elements and 31 contained more than one vocal element (e.g. "yeah yeah"), but this study does not distinguish between different vocalisations ('mhm', 'yeah', 'right', etc) and includes both clarification requests and sentence completions as backchannel responses (as discussed previously, we consider all these to be types of feedback—although this does not mean they will occur in the same contexts). Importantly for our proposed model, 27% of mhms (vocal backchannels) and 33% of nods were what Duncan (1974) terms 'early', or "prior to the end of the unit". For us here, it is what happens *after* this early feedback that would count as evidence for, or against our model.

More recently, Kjellmer (2009) investigated 6 common backchannel tokens, by sampling 1000 cases of each from the COBUILD corpus, discarding those he did not consider to be backchannels (e.g. which were not standalone, and/or which led to the producer of the backchannel taking the subsequent turn). Importantly for our analysis Kjellmer was primarily interested in the 42% of his sample that were 'turn-internal' backchannels—"cases where the speaker continues his/her turn across the backchannel". Of these, 71% occur at what, based on the syntax, he terms low interference places: "The places where the insertion of backchannels can be expected to occasion little interference, or none at all, i.e. places where a thought unit has been presented and where it is natural to pause," (e.g. between clauses) which could be analogous to the notion of FRSs as presented above. A further 10% occur at places "where a moderate degree of interference is possible [which] are also places where pauses naturally occur" and 16% at places where there is a "high degree of potential interference".

---

[6] We thank one of our reviewers, Mark Dingemanse, for helping us clarify this point.

These cases are particularly interesting for us, as they might be considered to be those which do not fit our model of FRSs (Howes and Eshghi 2017), but as we shall see, some of the features of DS allow us to explain these cases in a way that formalises the intuitions in Kjellmer (2009) about examples in which the backchannel occurs early due to the predictability of what follows—see Sect. 6.2. Note also that this proportion is in line with the 'early' numbers from Duncan (1974), as described above.

These findings have parallels with corpus studies of compound contributions in which the majority (63%) of same-speaker compound contributions are separated only by a backchannel, with 29% of same-speaker ones occurring after the first part of the turn did not end in a complete way (Howes et al. 2012a). However, both Kjellmer (2009) and Duncan (1974) have a more nuanced notion of a turn (using e.g. prosody/intonation), rather than potential turn completeness as based on transcriptions alone.

Neither Duncan nor Kjellmer, however, considers overlap, but there is evidence that a high proportion of backchannels do occur in overlap. Rühlemann (2007), for example, located standalone occurrences of the the two most common feedback indicators 'yeah' and 'mm' in the spoken BNC, finding that 17% and 10% of these occur in overlap respectively. Rühlemann (2007) takes this as evidence of backchannels supportive nature (in line with the 'continuer' function of Schegloff 1982) "backchannelling can be characterized essentially as 'non-turn-claiming talk' " (which characterisation goes back to Yngve (1970)—the originator of the term backchannel) but raises the question as to whether the overlapping backchannels are more likely to be those which do not occur at FRSs (those which are 'early' or 'high-interference' in Duncan's and Kjellmer's terminology). Note that due to the methodology of selecting all standalone occurrences of 'yeah' and 'mm', this study will miss cases which are not standalone (e.g. "yeah yeah"), and will include cases which are not usually considered to be backchannels (though see Fujimoto 2007), such as affirmative answers to yes/no questions, which may be less likely to occur in overlap, so these proportions are likely to be underestimates (as Rühlemann acknowledges).

Other corpus studies that cover aspects of feedback include Fernández (2006), whose annotations of non-sentential utterances (NSUs) in a subcorpus of the BNC include the classes 'acknowledgements' (5% of all utterances), and 'clarification ellipsis' (1%). However, as her focus is on NSUs, Fernández (2006) excludes cases in overlap, which as we have seen (Rühlemann 2007) means many genuine feedback utterances will be missed. For clarification requests, the numbers reported in Fernández (2006) are also an underestimate, as she is not concerned with sentential cases (e.g. "what do you mean?"). In another BNC study, Purver (2004) found that clarification requests made up just under 3% of utterances.

2.4 Dynamic Syntax Model of FRSs

According to the model presented in Howes and Eshghi (2017), FRSs can be defined as points during processing where a semantic sub-tree has just been completed. In practice this means that all TRPs and clause boundaries are FRSs, for example, but FRSs also occur after e.g. a complete noun phrase has been processed. Feedback

should not, however, normally be licensed e.g. within a noun or prepositional phrase, such as after a determiner but before the noun has been encountered. Feedback is therefore licensed only after a complete semantic unit of information has been parsed.

Before presenting the model from Howes and Eshghi (2017) in detail in Sect. 3.4, we first outline the formal tools of Dynamic Syntax (Kempson et al. 2001, 2016) and Type Theory with Records (Cooper and Ginzburg 2015; Cooper 2012, 2005) on which the model is based.

3 Dynamic Syntax

Dynamic Syntax (DS, Kempson et al. 2016; Cann et al. 2005; Kempson et al. 2001) is an action-based grammar framework that directly captures the time-linear, incremental nature of the dual processes of linguistic comprehension and production, on a word by word or token by token basis.[7] It models the linear construction of *semantic* representations (i.e. *interpretations*) as progressively more linguistic input is parsed or generated. DS is idiosyncratic in that it does not assume, or recognise, an independent level of syntactic representation over words: syntax on this view is sets of constraints on the incremental processing of semantic information in potentially multiple modalities (e.g. language and vision).

The output of parsing any given string of words, or non-verbal tokens, is thus a *semantic* tree representing its predicate-argument structure. DS trees are always binary branching, with argument nodes conventionally on the right and functor nodes to the left; tree nodes correspond to terms in the lambda calculus, decorated with labels expressing their semantic type (e.g. $Ty(e)$) and logical formulae—here as record types of Type Theory with Records (TTR, see Sect. 3.2 below); and beta-reduction determines the type and formula at a mother node from those at its daughters (Fig. 2). These trees can be *partial*, containing unsatisfied *requirements* potentially for any element (e.g. $?Ty(e)$, a requirement for future development to $Ty(e)$), and contain a *pointer*, $\diamondsuit$, labelling the node currently under development.

Grammaticality is defined as parsability in a context: the successful incremental word-by-word construction of a tree with no outstanding requirements (a *complete* tree) using all information given by the words in a string. We can also distinguish *potential grammaticality* (a successful sequence of steps up to a given point, although the tree is not complete and may have outstanding requirements) from *ungrammaticality* (no possible sequence of steps up to a given point) (Cann et al. 2007).

3.1 Actions in DS

The parsing process, which is what constitutes 'syntax' in DS, is defined in terms of conditional *actions*: procedural specifications for monotonic semantic tree update. *Computational actions* are general structure-building principles; and *lexical actions* are language-specific actions corresponding to and triggered by specific lexical tokens.
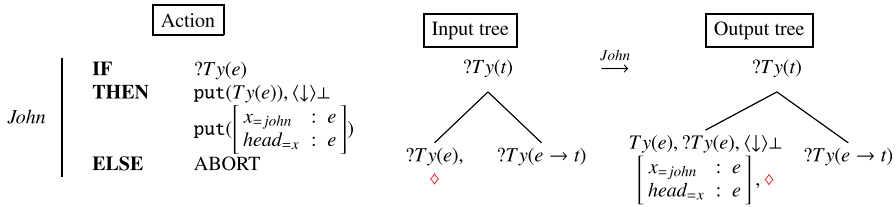

Fig. 1 Lexical action for the word 'John'

All actions take the form of 'macros' to provide update operations on semantic trees, instantiated as IF…THEN…ELSE rules which yield semantically transparent structures when applied (see e.g. Figs. 1 and 6).

Computational Actions form a small, fixed set of macros. Some encode the properties of the lambda calculus and the logical tree formalism (the logic of finite trees; LOFT: Blackburn and Meyer-Viol 1994) e.g. THINNING, which removes satisfied requirements; and ELIMINATION, which performs beta-reduction of a node's daughters, and annotates the mother node with the resulting formula. Other computational actions reflect the fundamental predictivity and dynamics of DS, e.g. COMPLETION, which moves the pointer up and out of a sub-tree once all requirements therein are satisfied; and ANTICIPATION which moves the pointer from a mother node to a daughter node with any unfulfilled requirements. While the former set of actions are *inferential*, thus not adding any new information to the trees, the latter set introduce alternative parse paths, thus capturing structural ambiguity: COMPLETION for example, precludes any further development of the current sub-tree because it moves the pointer up and out of it. In general, computational actions can apply optionally whenever their preconditions are met, but are not triggered by lexical input. The successful parse of a word w_1 thus amounts to finding a sequence of computational actions (possibly empty) that leads to a tree which satisfies the preconditions of the lexical action for w_1. The parse search process/history can be represented as a Directed Acyclic Graph (DAG), with (partial) semantic trees as nodes, and actions as edges, i.e. transitions between trees.

Lexical actions are associated with word forms in a DS lexicon. Like computational actions, these are tree-update macros composed of sequences of atomic actions which are basic tree-building operations such as `make`, `put` and `go`. Make creates a new (daughter) node, `go` moves the pointer, and `put` decorates the pointed node with a node label. Figure 1 shows an example for a proper noun, *John*. The action checks whether the pointed node (marked as ◇) has a requirement for type *e*; if so, it decorates it with type *e* (thus satisfying the requirement), the semantics for *John* (see Sect. 3.2 for details) and the bottom restriction ⟨ ↓ ⟩⊥ (meaning that the node cannot have any daughters). Otherwise (if there is no requirement $?Ty(e)$), the action aborts, meaning that the word *'John'* cannot be parsed in the context of the current tree.
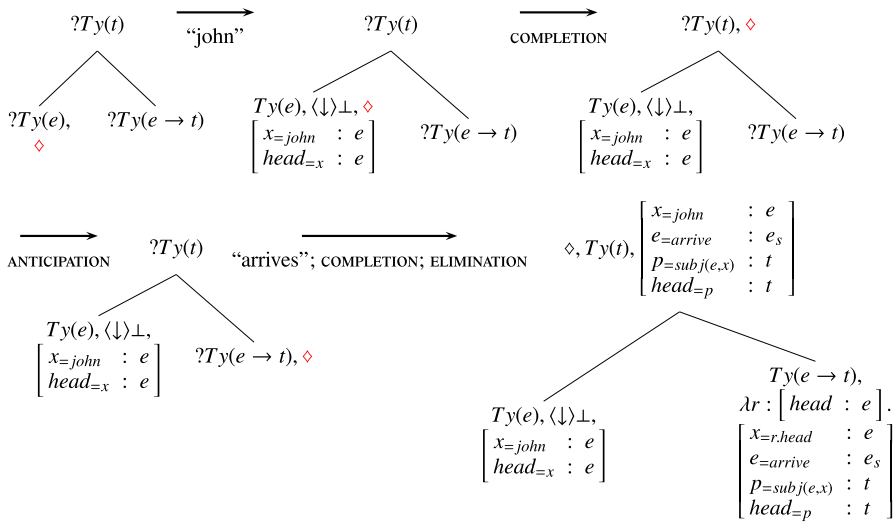
Fig. 2 Incremental parsing in DS-TTR: *"John arrives"*

Figure 2 shows "John arrives", parsed incrementally, starting with an empty tree, with only the root node's daughters created, and ending with a complete tree. The intermediate steps show the effects of COMPLETION, which moves the pointer up and out of a complete node—this process is central in our explanation of FRSs; and of ANTICIPATION, which moves the pointer down from the root to its functor daughter.

3.2 Type Theory with Records (TTR)

Recent efforts have incorporated TTR (Cooper 2012, 2005) as the semantic formalism in which meaning representations are couched in DS (Eshghi et al. 2012; Purver et al. 2011, 2010)—it is within the DS-TTR fusion that we express our model.

TTR is an extension of standard type theory, and has been shown to be useful in contextual and semantic modelling in dialogue (see e.g. Ginzburg 2012; Fernández 2006; Purver et al. 2010, among many others), as well as the integration of perceptual and linguistic semantics (Larsson 2015; Dobnik et al. 2012; Yu et al. 2016). With its rich notions of underspecification and subtyping, TTR has proved crucial for DS research in the incremental specification of content (Purver et al. 2011; Hough 2015); specification of a richer notion of dialogue context (Purver et al. 2010); models of DS grammar learning (Eshghi et al. 2013a, 2012); and models for learning dialogue systems from data (Eshghi et al. 2017; Kalatzis et al. 2016; Eshghi and Lemon 2014).

In TTR, logical forms are specified as *record types*, which are sequences of *fields* of the form [*l* : *T*] containing a label *l* and a type *T*. Record types can be witnessed (i.e. judged true) by *records* of that type, where a record is a sequence of label-value pairs [*l* = *v*]. We say that [*l* = *v*] is of type [*l* : *T*] just in case *v* is of type *T*.

Fig. 3 Example TTR record types

$$R_1 : \begin{bmatrix} l_1 & : & T_1 \\ l_{2=a} & : & T_2 \\ l_{3=p(l_2)} & : & T_3 \end{bmatrix} \quad R_2 : \begin{bmatrix} l_1 & : & T_1 \\ l_2 & : & T_{2'} \end{bmatrix} \quad R_3 : []$$

Fields can be *manifest*, i.e. given a singleton type e.g. $[l : T_a]$ where T_a is the type of which only a is a member; here, we write this as $[l_{=a} : T]$. Fields can also be *dependent* on fields preceding them (i.e. higher) in the record type (see Fig. 3).

The standard subtype relation $\sqsubseteq$ can be defined for record types: $R_1 \sqsubseteq R_2$ if for all fields $[l : T_2]$ in R_2, R_1 contains $[l : T_1]$ where $T_1 \sqsubseteq T_2$. In Fig. 3, $R_1 \sqsubseteq R_2$ if $T_2 \sqsubseteq T_{2'}$, and both R_1 and R_2 are subtypes of R_3. This subtyping relation allows semantic information to be incrementally specified, i.e. record types can be indefinitely extended with more information and/or constraints.

3.3 Dynamic Syntax and Dialogue Modelling

Given the inherent properties of the DS formalism, it has lent itself particularly well to dialogue modelling and analysis, and this has been a focus of research in the past decade or so (see Purver et al. 2006; Gargett et al. 2009; Gregoromichelaki et al. 2011; Howes 2012; Eshghi et al. 2015; Kempson et al. 2016, among many others). In DS, dialogue is modelled as the interactive and incremental construction of contextual and semantic representations (Eshghi et al. 2015). The contextual representations afforded by DS are of the fine-grained semantic content that is jointly constructed by interlocutors (see below for details), as a result of processing questions and answers, mid-utterance self-corrections (Hough and Purver 2012), various types of cross-turn context-dependency and ellipsis (Kempson et al. 2015); and split utterances (Howes et al. 2011; Howes 2012; Kempson et al. 2016).

3.4 Processing Feedback in Dynamic Syntax

In DS, *context*, required for processing various forms of context-dependency—including pronouns, VP-ellipsis, self-repair and short answers—is the parse search Directed Acyclic Graph (DAG). Edges correspond to DS actions—both Computational and Lexical actions—and nodes correspond to semantic trees after the application of each action (Sato 2011; Eshghi et al. 2012; Kempson et al. 2015)—see Fig. 4. Here, we take a coarser-grained view of the DAG with edges corresponding to words (sequences of computational actions followed by a single lexical action) rather than single actions, and dropping abandoned parse paths (see Hough 2015, for details)—Fig. 5 shows an example.

As Eshghi et al. (2015) show, grounding (the integration into the context of feedback) in a dyadic dialogue can be captured using the context DAG, augmented with two *coordination pointers*: the *self-pointer*, ♦; and the *other-pointer*, ◇, marking the points up to which the dialogue participants have each grounded the material. We dub this augmented context DAG, the *Interaction Control State* (henceforth ICS).
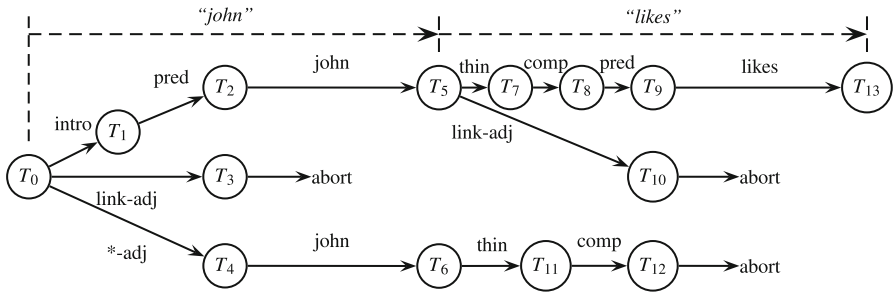
"john"　　　　　　　　　　　　　　　　　　　"likes"

pred T_2　john　T_5　thin　comp　T_7　T_8　pred　T_9　likes　T_{13}

intro T_1

T_0　T_3　abort

link-adj　link-adj　T_{10}　abort

*-adj　T_4　john　T_6　thin　T_{11}　comp　T_{12}　abort

Fig. 4 DS parsing as a directed acyclic graph (DAG): actions (edges) are transitions between partial trees (nodes)

Utterance	Context-final Semantics	A's Context After Utterance

A: The doctor

$$\left[r \quad : \left[\begin{array}{ll} x & : e \\ p = doctor(x) & : t \end{array} \right] \\ x_{=\iota(r.x,r)} \; : \; e \right]$$

$\diamond$ *The*　*doctor* $\blacklozenge$
$s_0 \rightarrow s_1 \rightarrow s_2$

B: mhm

$$\left[r \quad : \left[\begin{array}{ll} x & : e \\ p = doctor(x) & : t \end{array} \right] \\ x_{=\iota(r.x,r)} \; : \; e \right]$$

The　*doctor* $\diamond\blacklozenge$
$s_0 \rightarrow s_1 \rightarrow s_2$

A: he examined me

$$\left[\begin{array}{ll} r & : \left[\begin{array}{ll} x & : e \\ p = doctor(x) & : t \end{array} \right] \\ x_{=\iota(r.x,r)} & : e \\ x1_{=spkr} & : e \\ ev_{=examine} & : es \\ p_{=subj(ev,x)} & : t \\ p1_{=obj(ev,x1)} & : t \end{array} \right]$$

The　*doctor* $\diamond$　*he*　*examined*　*me* $\blacklozenge$
$s_0 \rightarrow s_1 \rightarrow s_2 \rightarrow s_3 \rightarrow s_4 \rightarrow s_5$

Fig. 5 Backchannels as movement of coordination pointers on interaction control states (ICS); from A's perspective

Any utterance causes ICS pointer movement, and interlocutors each have their own ICS which, as we will see below, can diverge at times, and re-converge as a result of clarification interaction and repair processes more generally. The self-pointer, $\blacklozenge$, on participant A's ICS tracks the point to which A has given evidence for reaching. The other-pointer, $\diamond$, tracks where the other participant, B, has given evidence for reaching. For example, an utterance produced by A will move A's self-pointer on their own ICS to the rightmost node of their ICS; on B's ICS, it is the other-pointer that moves to the same location. On this model, the intersection of the path back to the ICS root from the self- and other-pointers is taken to be grounded, with the effect that parse or production search within this grounded pathway is precluded, thus removing the computational cost associated with finding alternative interpretation pathways, as well as formally explaining how conversations move forward.

This model has been shown to account for backchannels, clarification interaction and other corrections (Eshghi et al. 2015). CRs cause branching on the ICS, where the

current path is abandoned and another branch constructed—a subsequent response to the CR plus the acknowledgement of this response eventually realigns the two coordination pointers, and the interlocutors' individual ICSs as a consequence—for reasons of space, here we suppress the details of how CRs are processed and integrated in the model, see Eshghi et al. (2015), Section 3.2, for a detailed exposition. By contrast, backchannels and utterance continuations do not create new branches, but move the other-pointer forward on the current path.

Figure 5 is a step by step illustration of how A's ICS develops as the dialogue proceeds, and as B's backchannel, 'mhm' is processed. After producing the first utterance, A's self-pointer, ♦, is on s_2, the right-most node of her ICS so far. B's backchannel "passes the opportunity to repair" (Schegloff 1982), thus moving A's other-pointer, ◇, to the same node and so grounding "the doctor". A's subsequent continuation creates new edges, and moves her self-pointer to the new right-most node. At this point, A's new utterance needs further feedback from B to be grounded: divergence of pointer positions thus represents 'forward momentum' in conversation (elsewhere called *discursive potential*; Ginzburg 2012).

This puts structural, surface forms of context-dependency at the centre of the explanation of participant coordination and feedback in dialogue: various forms of context-dependent expression, from the weakest—backchannels, which have little or no semantic content, to the strongest—utterance continuations, all serve to narrow down the otherwise mushrooming space of interpretation pathways. Their pervasiveness is therefore not coincidental, but strategic, and serves to make interpretation in dialogue *locally* computationally tractable.

This account gives formal rigour to the view expressed above under which language provides a set of interactional mechanisms—such as ellipsis, repair, and, indeed, backchannels—for dealing with the persistent potential for miscommunication (Healey et al. 2018; Kempson et al. 2016). So backchannels on this view really are "betokening the absence of such [communication] problems" (Schegloff 1982).

3.5 Modelling Feedback Relevance Spaces

Following Howes and Eshghi (2017), we take feedback to be normally licensed when a complete semantic unit has been fully processed and compiled, i.e. that a complete (sub-)tree has been constructed with no more tree compilation actions (namely THINNING and ELIMINATION, see Sect. 3.1) possible. Conceptually, these (sub-)trees are semantic, as it is the process of incrementally building up interpretations which is the core of language use and the driver of the DS grammar-as-parser approach. It is at these points that there may be enough information for the hearer to know whether or not repair is required, though the need for repair may also only become apparent further downstream (I may, for example successfully parse a name at the start of an utterance, but be thinking of the wrong individual, which only becomes apparent later in the utterance when I try to integrate the information—if ever).[8]

---

[8] Note that other factors of the interaction process can also play a role here—for example, you may not understand a word or two, but wait for more input because you believe the overall gist may become apparent from the later context, or you may ignore an obvious miscommunication as not interactionally relevant (so-

mhm	IF	$\exists x.?x$			
	THEN	abort			
	ELSE	IF	$\langle\uparrow_0\downarrow_1\rangle\exists x.Tn(x)$		
			$\langle\uparrow_0\downarrow_1\rangle\neg\exists x.?x$		
		THEN	abort		
		ELSE	IF	$\langle\uparrow_1\downarrow_0\rangle\exists x.Tn(x)$	
				$\langle\uparrow_1\downarrow_0\rangle\neg\exists x.?x$	
			THEN	abort	
			ELSE	$go(\langle\uparrow\rangle)$	

Fig. 6 Lexical entry for a backchannel

Clarification requests are modelled in DS using the existing, independently motivated mechanism of LINK-ADJUNCTION, which also accounts for both non-restrictive and restrictive readings of relative clauses (Cann et al. 2005). This creates a LINKed tree off the root node of the sub-tree with a requirement for a copy of the semantics of the node that the LINKed tree is developed from. Without requiring any additions to the grammar (c.f. Ginzburg and Cooper 2004; Purver 2004), this analysis includes CR fragments such as as reprises (e.g. 'Bill?') and sluicing (e.g. 'who?'), which are also treated as adjuncts.[9] This mechanism involves backwards search for the trouble source, thus moving the DS tree pointer out of the current tree and to the antecedent sub-tree on the context DAG. This backward search is normally precluded or at least predicted to be more computationally expensive if the current sub-tree under development is incomplete, in line with experimental findings that mid-constituent (non-FRS) reprise fragment CRs are more disruptive than between-constituent (FRS) ones (Healey et al. 2011; Eshghi et al. 2010).

Given that we take backchannels to be passing up an opportunity for repair, it follows that they are also licensed when a complete semantic sub-tree has been compiled. Figure 6 shows the lexical entry for backchannels.

The action in Fig. 6 is successful only if we are on a complete node (with no outstanding requirements) whose sister node if any—referenced via the $\langle\uparrow_1\downarrow_0\rangle$ and $\langle\uparrow_0\downarrow_1\rangle$ modalities—is not also complete. This ensures that all possible sequences of tree compilation actions (namely THINNING and ELIMINATION, see Sect. 3.1) and any subsequent COMPLETION actions are carried out *before* the backchannel can be parsed. The effect on the parse DAG (see e.g. Fig. 4) is that all (otherwise available) branches other than the one ending in a *maximally complete* tree are pruned. In addition, since the action pushes the pointer up and out of the complete sub-tree, any adjunct on or further qualification of the complete semantic unit via LINK-ADJUNCTION is precluded. Backchannels, especially those at FRSs that are not also TRPs (such as those which ground a noun phrase subject) are thus a strong signal that there is no need for repair, which, in pruning the parse DAG mean that any subsequent repair of a previously grounded element should be disruptive. Variants of this mechanism have been tested in the DS parser implementation (Eshghi et al. 2011; Eshghi 2015).

---

called *accepted misconstruals* in Clark 1997). We leave aside such considerations here, as they are outside the remit of the DS model.

[9] For reasons of space we do not include the analysis here, but see Gargett et al. (2009) and Eshghi et al. (2015).

Together, the above DS-internal dynamics allow us to provide a parsimonious explanation of why feedback ought to occur at FRSs, i.e. immediately after a complete semantic unit of information, using nothing beyond the already existing mechanisms of DS.

4 Hypotheses

Given the model in Howes and Eshghi (2017), presented above, we make the following predictions:

1. Feedback should ordinarily occur at FRSs—that is, after a complete semantic unit of information according to the DS model.
2. Early feedback indicates that the speaker's truncated semantic unit was locally predictable enough to be already understood without having been fully articulated. This means that there should be no need for the speaker to complete their utterance, and we expect a greater proportion of abandoned utterances after early feedback.
3. Late feedback addresses an earlier semantic unit, and is thus more disruptive due to the backward and forward search required on the part of the speaker to integrate them, and resume their utterance. Self-repair is known to provide "a measure of the difficulty of producing a contribution" (Healey et al. 2005, p.125) and pilot experiments on CRs suggest extra effort should lead to more utterance initial self-repair following an interruption (Eshghi et al. 2010; Healey et al. 2011).
4. Acknowledgements are less disruptive than CRs, as they do not require any specific response from the speaker. They are therefore more likely to occur in overlap, and more likely to occur at non-FRSs than CRs.

5 Corpus Analysis

5.1 Materials

For this study we used 20 dialogues (made up of 4938 utterances) from the sub-corpus of the BNC that had already been annotated for Compound Contributions (Purver et al. 2009; Howes et al. 2012a) and non-sentential utterances (Fernández 2006). In particular, we make use of the `end-complete` tags from Purver et al. (2009) which indicates whether turns could be considered to be finished in a complete way or not. We term those utterances which do not end in a complete way and do not have a later continuation (according to the pre-existing annotations) 'abandoned utterances'. We also use the notion of 'repairs' from Purver et al. (2009), which we term 'restarts'. For our purposes, an utterance following a piece of feedback contains a restart if it begins with an utterance initial self-repair in the form of a repeated or paraphrased word or phrase (as in Example 6).

Our main annotation task was to annotate each utterance that was an acknowledgement or a CR (enabling us to include examples discounted by Fernández (2006) as discussed above). For `end-complete` tags, Purver et al. (2009) reported inter-annotator agreement using Cohen's Kappa of between 0.73 and 1 for three dialogues

Table 1 Annotation tags

Tag	Value	Explanation
End-complete	y/n	For all utterances: does this utterance end in such a way as to yield a complete proposition or speech act?
Ack/clarify	Ack/CR/(blank)	For all utterances: does this sentence consist of a backchannel (e.g. 'yeah', 'mhm', 'right') or a repeated word or phrase acknowledging the proposition or speech act of a previous utterance? If not, does this utterance consist of a clarification request, clarifying the proposition or speech act of a previous utterance?

annotated by three linguistically knowledgeable annotators. For acknowledgements and CRs, 9 naive coders (Masters students, for course credit) each annotated a dialogue which were then all annotated by one of the authors. Interrater reliability was moderate to high for all 9 dialogues $\kappa = 0.64 - 0.93$ (average 0.78). The annotations and instructions given to the annotators are publicly available at https://osf.io/8fjbh/.

5.2 Method

The high-level annotation task is as described in Table 1 (with end-complete tags already annotated, as above). Following this initial pass, we further examine those cases which occur after a non end-complete utterance or in overlap with an end-complete utterance to assess whether these in fact occur at what our DS model predicts to be an FRS, or if not whether they can be classified as appearing just after an FRS ('late') or before an FRS ('early'). Following Duncan (1974), we took there to be an intuitive notion of whether feedback occurred early or late. Of course, each piece of feedback that does not occur at an FRS could be related to the preceding (complete) semantic unit, or to the current (incomplete) semantic unit in progress.[10] In practice, we categorised feedback as late only if it was obviously taken as relating to the preceding semantic unit (see examples, below). All other cases, including some ambiguous cases that may have related to a prior semantic unit and thus properly been counted as late, we categorised as early, meaning that the proportions we report as late are likely to be an underestimate. Inter-rater reliability between the authors for the three way coding (at_FRS, early or late) of a sample of 96 cases which were not automatically coded as at an FRS (see footnote 10) was $\kappa = 0.71$.

5.3 Results

13.3% of utterances in our sample were acknowledgements (more than double that found by Fernández (2006), though as discussed above she did not include many cases for methodological reasons), and 2.3% were CRs (slightly lower than Purver (2004),

---

[10] Note that the 539 cases which occurred after a end-complete tag, and contained no overlap were automatically annotated as occurring at an FRS, leaving only 231 cases to be manually annotated as occurring at an FRS or early or late.

Table 2 Summary data

	Total		After non End-complete		After End-complete		Occurring in Overlap		No Overlap	
	#	%	#	%	#	%	#	%	#	%
Acknowledgements	655	13	112	17	543	83	174	27	481	73
CRs	115	2	10	9	105	91	14	12	101	88
Total	770	16	122	16	648	84	188	24	582	76
(Other utterances)	4168	84	228	5	3940	95	687	16	3481	84
Total utterances	4938		350	7	4588	93	875	18	4063	82

but this could be accounted for the fact that he found a higher proportion of CRs in the demographic—non-context specific—dialogues which constituted 90% of his sample, but represents only 5% of our sample). In all the statistical results reported below, particularly with regards to CRs, we must be cautious in our interpretations due to the small number of cases involved.

5.4 Positioning of Feedback

In support of hypothesis 1, 543 acknowledgements (83%) and 105 CRs (91%) occurred after utterances which could be considered end-complete. This means that the utterance ends in such a way as to yield a complete proposition or speech act, and so feedback in these cases occurs at turn or clause boundaries (though see below regarding cases which appear in overlap with an end-complete turn). As these are all potentially TRPs, they also correspond to places that are FRSs according to the model in Howes and Eshghi (2017).[11]

In terms of overlap, a greater proportion of acknowledgements (174/655: 27%) than CRs (14/115: 12%) occur in overlap with another speaker's talk ($\chi_1^2 = 10.978$, $p = 0.001$), supporting hypothesis 4. As expected, the proportion of acknowledgements in our study is greater than that in Rühlemann (2007). Acknowledgements are also significantly more likely than CRs to occur after a non end-complete utterance (112/655 = 17% vs 10/115 = 9%: $\chi_1^2 = 5.181$, $p = 0.02$).

Of those 112 acknowledgements which did not occur after a potentially end-complete utterance a higher proportion than expected by chance occur in overlap (77: 69%; $\chi_1^2 = 123.442$, $p < 0.001$) which is not the case for CRs (2: 20%), also offering support for hypothesis 4.

Closer examination of the feedback that occurs mid-utterance (i.e. after a non end-complete utterance or in overlap with an end-complete utterance, Table 3), shows that many of these do occur at what our model predicts to be an FRS (ack: 109/209, 52%; CR: 6/22, 27%) and which correspond to Kjellmer's (2009) "low-interference places", see e.g. Examples 2–3. This is especially the case when we consider those cases with some degree of overlap (Example 4)—due to the transcription conventions of the BNC, overlapping talk may be shown interleaved with the preceding utterance, but note that examples such as this one could equally be transcribed in a way that makes it obvious that they do occur at FRSs (see the alternate transcription in Example 5).

When we manually annotate the cases described in footnote 10, 115 (of 231) instances were assessed as occurring at FRSs. The difference between the total proportion of acknowledgements and CRs that occur at non-FRSs (100/655, 15%; 16/115, 14%) is therefore not significant (see summary in Table 4), contra hypothesis 4.

---

[11] Note that as possible end-completeness was annotated purely based on the transcriptions this will inevitably include some feedback which would be considered mid-turn by Duncan (1974) and Kjellmer (2009).

Table 3 Positioning of mid utterance feedback

	After non end-complete utterance							Overlapping end-complete utterance						
	Total	Early		At FRS		Late		Total	Early		At FRS		Late	
	#	#	%	#	%	#	%	#	#	%	#	%	#	%
Acknowledgements	112	46	41	52	46	14	13	97	35	36	57	59	5	5
CRs	10	3	30	2	20	5	50	12	4	33	4	33	4	33
Total	122	49	40	54	44	19	16	109	40	37	61	56	8	7

Table 4 Summary: positioning of feedback

	Total	Early		At FRS		Late	
	#	#	%	#	%	#	%
Acknowledgements	655	81	*12*	555	*85*	19	*3*
CRs	115	7	*6*	99	*86*	9	*8*
Total	770	88	*11*	654	*85*	28	*4*

Example 3 BNC file HDK:162–164—within-turn acknowledgement at an FRS

I   162   No, I, I wondered because er it was quite an honour to have your name

A   163   Yes.

I   164   put on the honour board.

Example 4 BNC file J8D:51–54—within-turn acknowledgement at an FRS, in overlap

A   51   As if we haven't got enough on our plate!

      52   [The thing]

U   53   [Mm]

A   54   that is stupid Terry is key stage four!

Example 5 BNC file J8D:51–54 (alternate transcription of Example 4)

A   51   As if we haven't got enough on our plate!

U   53   [Mm]

A   54   [The thing] that is stupid Terry is key stage four!

5.5 Late Feedback

As shown in Table 4, of the 100 acknowledgements that do not occur at an FRS, 19 (19%) are categorised as late, as are 9 of the 16 CRs (56%, a significantly higher proportion $\chi^2_1 = 10.452$, $p = 0.001$). In the case of both acknowledgements and CRs, these occur soon after an FRS—often after a conjunction ('and', 'or', or 'but' in 11 of the 28 'late' cases—8/19 Acks (42%); 3/9 CRs (33%)) or new clause beginning (e.g. 'as', 'if', 'that' or 'so', or a subject noun phrase in 11 cases 7/19 Acks (36%); 4/9 CRs (44%)), and the feedback is clearly interpreted as applying to the preceding (end-complete) segment, as shown in Examples 6 and 7.

Example 6 BNC file G4W:144–146—late acknowledgement, after a conjunction

D	144	I'm not saying that the same person thought of these two things at once, but
U	145	Okay.
D	146	these are these are things which were all said.

Example 7 BNC file G4W:136–139—late CR, after a relative clause marker

D	136	Er women were also said to be polite, diffident, verbose and deferential, which
U	137	What all of those?
D	138	Mhm.
U	139	⟨unclear⟩ polite and deferential.

In support of hypothesis 3, restarts (utterance initial self-repairs), as seen in Example 6, are more common following late feedback than feedback at an FRS (7/28 = 25% vs 66/654 = 10% $\chi_1^2 = 6.244$, $p = 0.012$). There is also a trend such that there are more restarts following late rather than early feedback, but this does not reach significance due to the low numbers of instances (7/28 = 25% vs 10/88 = 11% $\chi_1^2 = 3.158$, $p = 0.076$). Note that even in the late cases, there is no utterance initial self-repair in the majority of cases.

In support of hypothesis 4, there also seems to be more disruption caused by late CRs than late acknowledgements. Although there is no statistical difference between the amount that have utterance initial repair (3/9 (33%) of the late CR cases and 4/19 (21%) of the late acknowledgement cases), there is a difference in the number of utterances which get abandoned (i.e. the interrupted utterance in progress never gets continued)[12] after the interlocutor produces a late CR compared to a late Acknowledgement (6/9 = 67% vs 1/19 = 5% $\chi_1^2 = 12.280$, $p < 0.001$). It is telling that all the 3 cases of late CRs which did not get abandoned, such as Example 8, included overlap with the CR.

Example 8 BNC file HDD:436–439—late CR

K	436	That's no good because the mass is on Thursday.
	437	It's [Thursday]
U	438	[Which Thursday]? ⟨whispered⟩
K	439	this Thursday coming.

---

[12] This is based on the `continues` annotation tag from Purver et al. (2009).

5.6 Early Feedback

More interesting are the 81 acknowledgements and 7 CRs that were categorised as early. In these cases, it appears that the feedback is placed within the semantic unit to which it relates, meaning that the listener has made some prediction about the upcoming linguistic material, and is able to produce feedback before the input is complete. This matches observations from both Duncan and Kjellmer: "…an early back channel may not be merely misplaced, but rather it may carry significant information for the interaction …[and] may indicate, not only that the auditor is following the speaker's message, but also that the auditor is actually ahead of it" (Duncan 1974); "the point at which the backchannel is inserted may suggest when the listener has formed an idea of what the speaker is conveying or going to convey […] the information-content of an utterance is often conveyed to the listener before it has been expressed *in toto*" (Kjellmer 2009, italics original).

This holds of both early backchannels and early CRs: in contexts where upcoming material is highly predictable, early CRs are likely to be formulated as a compound contribution (Howes et al. 2012b), directly continuing the truncated utterance, and, as opposed to early backchannels, actually *articulating* the predicted material as a clarification question. We discuss these observations in detail below.

Examination of the data shows that early feedback often occurs at points at which the remainder of the semantic information unit is already predictable, with the context of the utterance constraining how it is likely to progress. In Example 9, the source of this predictivity is the extremely local context of a specific compound noun phrase "North York Moors", which is familiar enough to the interlocutors that the listener can understand what is being talked about (and therefore produce an acknowledgement) before the complete noun phrase has been uttered. In Example 10, the source of the predictability is also the local context, but in this case it is due to the repetition of the form of the lexical items.

Example 9 BNC file G3Y:180–184—predictive acknowledgement

M	180	I can't stand the seaside.
	181	Couldn't we go to the North York [Moors]
L	182	[Mm.]
M	183	instead.
	184	Right so instantaneous ⟨unclear⟩ .

Example 10 BNC file G3Y:69–71—predictive acknowledgement

M	69	And erm you work out whether your preference is very strongly this way or a bit this way or a bit that [way].
L	70	[Mm].
M	71	And just put a mark on the line to indicate the strength of the preference.

In contrast, in Examples 11 and 12 the source of the predictability of the utterance in progress at the point at which the feedback is provided is more general; the interlocutors are both aware of the shared conversational history, and they are talking about something that they have already been talking about.[13]

Example 11 BNC file KBG:35–38—predictive acknowledgement

C	35	No, you were right what you said though about
S	36	Mm.
C	37	Matthew.
S	38	Yeah I just ⟨pause⟩ it just seems to me such a complete ⟨pause⟩ and utter waste of of, his time …

Example 12 BNC file KBG:150–152—predictive acknowledgement, overlap

S	150	Mind you, there's been a great pause on it while you went to see to the [baby].
C	151	[Mm].
	152	So then they're all the details, and there's examples and

These factors can overlap, with the utterance continuation constrained both by the lexical items, and the broader context of the topic being discussed, as in Example 13.

Example 13 BNC file G3Y:307–312—predictive acknowledgement

M	307	And the thinking people prefer to use impersonal objective material.
L	308	Mm.
M	309	Whereas the feeling people prefer subjective personal
L	310	[Mm].
M	311	[material].
	312	So you've go got a situation where if somebody wants to change something they're actually, you know …

Interestingly, this predictability of upcoming content means that the utterance in progress when the early backchannel occurs can in fact be abandoned, with the assumption that both participants know roughly what would have been said. Utterances can therefore be interpreted as if they were complete—even if they are never fully articulated, as in Example 14 (in which M is getting L to fill out a personality test—in this example, M is checking whether L has decided where to rate herself along the second scale) and Example 15. In Example 14 the source of the predictability may also come from another modality, for example, L might have demonstrated her understanding

---

[13] Note that Example 11 could have been a late acknowledgement that ought to have occurred after 'though' but before 'about', however the next turn suggests that it was taken as acknowledging 'Matthew'.

of M's incomplete question by marking her answer on paper.[14] For DS, these examples mean that incomplete syntactic strings can, in particular interactional contexts, result in complete semantic interpretations. We discuss how this might be modelled in Sect. 6.2. In our data, although numbers are too small for tests of statistical significance, a higher proportion of utterances are abandoned (i.e. never completed) after early than late acknowledgements (11/81: 14% vs 1/19: 5%) in line with hypothesis 2. In contrast, utterances in progress are less likely to be abandoned after an early rather than a late CR (1/7: 14% vs 6/9: 67%). This seems to be because of the work required in order to address a CR, as shown in Example 7, which is not present in the case of an acknowledgement.

Example 14 BNC file G3Y:286–288—predictive acknowledgement, utterance never completed

M	286	Do you do you have a
L	287	Yeah.
M	288	Okay, right.

Example 15 BNC file JN7:66–70—predictive acknowledgement, utterance never completed

S	66	Yeah ⟨pause⟩ Well it's it's it's a it's a bit of a weighty subject that.
	67	I think we ought to er
B	68	Yeah okay
S	69	Why don't you and I talk about it separately then
B	70	Yeah alright

Early CRs on the other hand are often formulated as continuations of the prior utterance (4/7 = 57%), as in Examples 16 and 17, and in fact this strategy has been shown experimentally to be more likely in contexts where the upcoming part of speech is highly predictable (Howes et al. 2012b). This is in line with our observation above that early backchannels appear to be placed within a highly predictable semantic unit, and treat the interrupted sentence as if already complete. The difference with CRs is that the predicted material is actually *articulated*. We will discuss this and the similarities to non-clarificatory compound contributions (Purver et al. 2011; Howes 2012; Kempson et al. 2016, a.o.) in terms of the DS model in Sect. 6.2.

---

[14] Such examples show that models of language use need to be integrated as part of a general model of action/perception and that natural language syntax has to be definable as constraints over the process of real-time, semantic growth—see Kempson et al. (2016) and Gregoromichelaki et al. (2020) for discussion of these fundamental requirements.

Example 16 BNC file G4X:185–191—predictive CR

S	185	But it's it's such a broad issue, that any opportunity we do get, to have a a broader discussion, for more individuals to contribute their or have their say, you know again on a workshop or something, I would very much support that.
C	186	We had actually got that in for that September
A	187	Mm, ⟨unclear⟩
C	188	on, we got a date for that on the C Os [day]
S	189	[Workshop?]
C	190	Yeah.
S	191	Oh good

Example 17 BNC file K69:109–112—predictive CR

J	109	How does it generate?
M	110	It's generated with a handle and
J	111	Wound round?
M	112	Yes, wind them round and this should, should generate a charge …

6 Accounting for the Data in the Model

In the rest of this paper, we focus on how feedback in the form of acknowledgements and clarification requests are processed and integrated, and how the model can account for their temporal distribution and licensing.

In light of our corpus study, we can see how the model accounts for the 85% of cases where feedback occurs at FRSs. We now look in more detail at what the model predicts should happen for late and early feedback, how these can be interpreted, and how well these predictions are borne out by the corpus data.

6.1 Late Feedback

In our model, there are two possibilities for when feedback is produced at a point where it is not licensed. The first is that the listener is lagging behind the speaker and has produced the feedback late. This may reflect the time taken for the listener to integrate the information into their interpretation—'correct' placement of feedback at FRSs will require some element of prediction, analogously to how turn-taking occurs with such precise timing indicating that people predict upcoming TRPs (de Ruiter et al. 2006, a.o.). In this case, feedback can be interpreted as grounding the most recent increment (informational unit).

For backchannels this means backtracking on the Interaction Control State (ICS, see Sect. 3.4) to the first, most local point at which the backchannel is parsable, or equivalently, where COMPLETION just occurred: this process involves backwards search, and is thus computationally expensive. The model therefore predicts that it should cause some disruption for the speaker who integrates the feedback. This is

consistent with our corpus results above (Sect. 5.5): restarts in the speaker's utterance are significantly more likely to occur after late feedback than they are after feedback at an FRS. Interestingly, despite this need to backtrack, there is often no visible sign of any disruption following a late backchannel. This may be because such disruption is at the level of timing in the form of e.g. short pauses, or elongations, which are not recorded in the transcripts. It may also be due to the way we have defined late feedback, which means that it is often only one word which needs to be retained in memory while the pruning backchannel operation is undertaken. In terms of the ICS, the backchannel action does not open up any new paths, so it is relatively simple to forward-track to the right-most point of the single open ICS parse path.

In contrast, late clarification requests, which require a response, are more disruptive. In these cases, the need to backtrack to the trouble-source antecedent of the CR often leads to the ICS path under development being *abandoned*. This is unsurprising in terms of the DS model as the parser has to both backtrack and create a new DAG path to resolve the CR (for details, see Eshghi et al. (2015), Fig. 6). This means that the interrupted (non-grounded part of the) utterance is both more distant in time and in terms of ICS pointer position, which may also reflect well-studied short-term memory limitations (Gibson 1991). Furthermore, the return to the point of interruption involves, unlike for backchannels, forward *search*, thus rendering it even more computationally expensive.

6.2 Early Feedback

The second, perhaps more interesting, case is where feedback is produced early. In this case, feedback seemingly precedes the completion of a semantic unit, which is not licensed by the grammar model.

Our proposal here is that the producer of an early feedback token has anticipated *the content* of the speaker's utterance before it is complete. We argue that this is potentially a computationally inexpensive process due to: (1) predictivity of the grammar system itself in providing the hearer with specific local requirements to be satisfied, e.g. a $?Ty(cn)$ which predicts a noun; (2) the parity of parsing and production mechanisms; and (3) contextual predictability, e.g. due to prior common ground among the interlocutors, with the incremental content so far having been enough for the hearer to predict the rest. The proposal is thus that the hearer has covertly projected content on the type incomplete node, even before the speaker has uttered the next word, and thus that the hearer's Interaction Control State (see Sect. 3.4 and Fig. 5) is such that they can produce the backchannel or clarification request early. The result is a realignment of the interactants' Interaction Control States albeit covertly.

In the case of clarification requests, this prediction means that the CR is often formulated as a direct continuation of the utterance in progress, as shown in Example 18—i.e. while it has the function of a CR, it takes the structural form of a compound contribution (Purver et al. 2009; Howes 2012), with questioning intonation. In Purver's 2004 terminology, these are *gap fillers* which are different from other types of CRs in that they do "not clarify a part of the original utterance as actually presented, but instead a part that was originally intended by the speaker but not produced" (Purver 2004, p. 68).

Example 18 BNC file KB2:5155–5159—Candidate clarification request predicted from context

A	5155	I says, ooh this gravy's lovely!
J	5156	Yeah!
A	5157	He says er, yeah he said I did some onion, and then, I got some of them, you know
J	5158	Granules?
A	5159	yeah, put some of that in
J	5160	Mm.

As shown in DS accounts of cross-person completions (Purver et al. 2010; Eshghi et al. 2012; Kempson et al. 2016), a listener may switch to being a speaker at any point in the interpretation of an utterance, provided that they have a more advanced goal tree in mind. This is precisely the mechanism being exploited in the early CR case, with the questioning intonation taken to indicate lack of confidence in the proposed continuation.

For early backchannels, we propose that the same predictive mechanisms are exploited, except that the predicted content is directly projected on the listener's as yet incomplete tree node while parsing, and that this is done *covertly* before the speaker has finished speaking; thus licensing the listener's early backchannel. In Example 19:211, for example, J's continuation is so predictable (it is a repetition of prior material; "got a lot on") that A does not have to wait for it in order to interpret the complete utterance (including the unuttered material that A has predicted will come next) but can instead rerun the actions she has already used, or project the requisite content on the as yet incomplete node directly.

Example 19 BNC file KB2:210–213 reuse of prior actions

J	210	her mum really she's got a lot on, she'll have a lot on cos she's got to prepare for that wedding, you know what you're like when you, [you've got]
A	211	[Mm]
J	212	you know if you want, want to be doing things [don't you get out of house and that]
A	213	[Yeah, pre- preparing for a wedding, yeah]

The difference with backchannels is that the hearer, instead of producing the actual completion of the speaker's incomplete turn, signals that there is no need for the speaker to carry on (14% of cases, see above), although of course, they may do so. Furthermore, projecting predicted content is a less computationally expensive process than realising the projected content with words because the latter involves lexical search.

This means that if you have a high degree of confidence that your projected continuation is correct, then an early backchannel is a good strategy. If, however, your projected continuation is more tentative, then producing the lexical items as a CR, to be confirmed or disconfirmed by your interlocutor is a better strategy to prevent possible misunderstandings further down the line, despite the locally higher cost.

This analysis is further supported by a text chat experiment in which turns were artificially truncated (Howes et al. 2012b). Some incomplete turns were responded

to as if they were complete, provided that the continuation was highly predictable. In addition, producing candidate completions as clarification requests was a fairly common strategy—particularly when the part of speech of the upcoming material was predictable, and the context sufficiently constrained.

7 Conclusions and Future Directions

We have presented a corpus study of feedback in British English dialogue and shown how it can be accounted for using Dynamic Syntax, which unifies the dialogue phenomena of backchannels, clarification requests and completions in terms of their potential for participant coordination in interaction. All these phenomena can occur subsententially, and serve as interactional mechanisms for the management of interlocutors' characteristic communicative divergence and convergence in dialogue. They achieve this essentially by local pruning of the search space of potential interpretation pathways, thus making this seemingly intractable space of possibilities tractable at a given point in an exchange.

Given the high occurrence of feedback at FRSs, our corpus study suggests that FRSs are interactionally relevant parts of a utterance, analogous to TRPs, and that these are based on low-level semantic criteria. Further, we provide an explanation of how, when feedback such as backchannels occurs at points other than FRSs it gets interpreted as if it had done so. This can be because the feedback is late and grounding the previous informational unit, or in specific interactional contexts because the feedback is early and the (rest of the) informational unit is predictable.

In this paper, our explanations of feedback at non-FRSs using the DS model have focused more on backchannels than CRs. This is partly due to the distributions of the types of feedback; there are simply not enough cases of early and late CRs to draw any strong conclusions. We are therefore running some more precisely controlled text chat experiments involving interruptive clarification requests, following the methodology of Healey et al. (2018).

The corpus study in this paper has focused on British English data. It is well-known that different languages have different frequencies and distributions of backchannels (see e.g. Kita and Ide 2007, on Japanese), for example, but it is not known whether these occur at more of the available FRSs or are influenced by other factors. Whatever the case may be, we would expect feedback positioning to interact directly with the grammar of the language. For example, the DS grammar of head final pro-drop languages such as Japanese (Kempson and Kiaer 2008) imposes different constraints on the unfolding utterance interpretation, in turn affecting FRSs. While we hope that the notion of FRSs is cross-linguistically applicable, this is, of course, an empirical question.

It also remains to be seen whether the model presented here or an extension of it can account for patterns of grounding and feedback in multi-person dialogue where participant role (Goffman 1981) has been shown to affect levels of semantic coordination and understanding (see Healey and Mills 2006; Eshghi 2009, a.o.), including situations in which more than one participant can form higher order units (dubbed *parties* by Schegloff 1995), in turn affecting feedback and grounding patterns (Eshghi and Healey 2015; Howes 2012).

As noted, this work has implications for the production and interpretation of feedback in artificial agents and interactive robots, specifically those intended to have human-like behaviours, e.g. companion robots for the elderly. Our study suggests that dialogue systems should produce or parse feedback not just based on unanalysed features such as prosody (which may result in accurate placement, but have no functional role in the dialogue system in terms of coordination), but because they have successfully compiled a semantic unit.

References

Bangerter, A., & Clark, H. H. (2003). Navigating joint projects with dialogue. *Cognitive Science*, *27*(2), 195–225.

Bavelas, J. B., Coates, L., Johnson, T., et al. (2000). Listeners as co-narrators. *Journal of Personality and Social Psychology*, *79*(6), 941–952.

Bavelas, J. B., De Jong, P., Korman, H., & Jordan, S. S. (2012). Beyond back-channels: A three-step model of grounding in face-to-face dialogue. In: *Proceedings of Interdisciplinary Workshop on Feedback Behaviors in Dialog* (pp. 5–6).

Blackburn, P., & Meyer-Viol, W. (1994). Linguistics, logic and finite trees. *Logic Journal of the Interest Group of Pure and Applied Logics*, *2*(1), 3–29.

Bolden, G. B., Hepburn, A., & Potter, J. (2019). Subversive completions: Turn-taking resources for commandeering the recipient's action in progress. *Research on Language and Social Interaction*, *52*(2), 144–158.

Bouzouita, M. (2008). At the syntax-pragmatics interface: Clitics in the history of Spanish, In R. Cooper & R. Kempson (Eds.), *Language in Flux: Dialogue Coordination* (pp. 221–264). London: Language Variation, Change and Evolution, College Publications.

Burnard, L. (2000). *Reference guide for the British National Corpus* (World ed.). Oxford: Oxford University Computing Services.

Cann, R., Kempson, R., & Marten, L. (2005). *The Dynamics of Language*. Oxford: Elsevier.

Cann, R., Kempson, R., & Purver, M. (2007). Context and well-formedness: The dynamics of ellipsis. *Research on Language and Computation*, *5*(3), 333–358.

Chatzikyriakidis, S., & Kempson, R. (2011). Standard Modern and Pontic Greek person restrictions: A feature-free dynamic account. *Journal of Greek Linguistics*, *11*, 127–166.

Chomsky, N. (1965). *Aspects of the Theory of Syntax*. Cambridge, MA: MIT.

Clark, H. H. (1996). *Using Language*. Cambridge: Cambridge University Press.

Clark, H. H. (1997). Dogmas of understanding. *Discourse Processes*, *23*(3), 567–598.

Cooper, R. (2005). Records and record types in semantic theory. *Journal of Logic and Computation*, *15*(2), 99–112.

Cooper, R. (2012). Type theory and semantics in flux. In R. Kempson, N. Asher, & T. Fernando (Eds.), *Philosophy of Linguistics, Handbook of the Philosophy of Science* (Vol. 14, pp. 271–323). Amsterdam: Elsevier, North Holland.

Cooper, R., & Ginzburg, J. (2015). *Type Theory with Records for natural language semantics* (Vol. 12, pp. 375–407). New York: Wiley. https://doi.org/10.1002/9781118882139.ch12.

de Ruiter, J. P., Mitterer, H., & Enfield, N. (2006). Projecting the end of a speaker's turn: A cognitive cornerstone of conversation. *Language*, *82*(3), 515–535.

Dobnik, S., Cooper, R., & Larsson, S. (2012). Modelling language, action, and perception in Type Theory with Records. In *Proceedings of the 7th International Workshop on Constraint Solving and Language Processing* (pp. 51–63).

Drew, P. (1991). Asymmetries of knowledge in conversational interactions. In K. Foppa & I. Markova (Eds.), *Asymmetries in Dialogue* (pp. 29–48). Hemel Hempstead: Wheatsheaf.

Duncan, S. (1972). Some signals and rules for taking speaking turns in conversations. *Journal of Personality and Social Psychology*, *23*(2), 283–292.

Duncan, S. (1974). On the structure of speaker–auditor interaction during speaking turns. *Language in Society*, *3*(2), 161–180.

Eshghi, A. (2009). Uncommon ground: The distribution of dialogue contexts. Ph.d thesis, Queen Mary University of London.

Eshghi, A. (2015). DS-TTR: An incremental, semantic, contextual parser for dialogue. In *Proceedings of the 19th SemDial Workshop on the Semantics and Pragmatics of Dialogue (goDial)* (pp. 172–173).

Eshghi, A., & Healey, P. G. T. (2015). Collective contexts in conversation: Grounding by proxy. *Cognitive Science*, *40*, 229–324.

Eshghi, A., & Lemon, O. (2014). How domain-general can we be? Learning incremental dialogue systems without dialogue acts. In *Proceedings of the 18th SemDial Workshop on the Semantics and Pragmatics of Dialogue (DialWatt)* (pp. 53–61).

Eshghi, A., Healey, P. G. T., Purver, M., Howes, C., Gregoromichelaki, E., & Kempson, R. (2010). Incremental turn processing in dialogue. In *Architectures and Mechanisms for Language Processing, York, UK*.

Eshghi, A., Purver, M., & Hough, J. (2011). DyLan: Parser for Dynamic Syntax. Technical report, Queen Mary University of London, EECSRR-11-05

Eshghi, A., Hough, J., Purver, M., Kempson, R., & Gregoromichelaki, E. (2012). Conversational interactions: Capturing dialogue dynamics. In S. Larsson & L. Borin (Eds.), *From Quantification to Conversation: Festschrift for Robin Cooper on the Occasion of his 65th Birthday, Tributes* (Vol. 19, pp. 325–349). London: College Publications.

Eshghi, A., Hough, J., & Purver, M. (2013a). Incremental grammar induction from child-directed dialogue utterances. In *Proceedings of the 4th Annual Workshop on Cognitive Modeling and Computational Linguistics (CMCL)* (pp. 94–103). ACL.

Eshghi, A., Purver, M., Hough, J., & Sato, Y. (2013b). Probabilistic grammar induction in an incremental semantic framework. In D. Duchier, Y. Parmentier (Eds.), Constraint Solving, and Language Processing. CSLP. (2012). Lecture Notes in Computer Science (Vol. 8114, pp. 92–107). Berlin, Heidelberg: Springer.

Eshghi, A., Howes, C., Gregoromichelaki, E., Hough, J., & Purver, M. (2015). Feedback in conversation as incremental semantic update. In *Proceedings of the 11th International Conference on Computational Semantics (IWCS)* (pp. 261–271). London: ACL.

Eshghi, A., Shalyminov, I., & Lemon, O. (2017). Bootstrapping incremental dialogue systems from minimal data: Linguistic knowledge or machine learning? In *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing (EMNLP)* (pp. 2220–2230).

Fernández, R. (2006). Non-sentential utterances in dialogue: Classification, resolution and use. Ph.d. thesis, King's College London, University of London.

Fujimoto, D. T. (2007). Listener responses in interaction: A case for abandoning the term, backchannel. *Journal of Osaka Jogakuin College*, *37*, 35–54.

Gargett, A., Gregoromichelaki, E., Kempson, R., Purver, M., & Sato, Y. (2009). Grammar resources for modelling dialogue dynamically. *Cognitive Neurodynamics*, *3*(4), 347–363.

Gibson, E. A. F. (1991). A computational theory of human linguistic processing: Memory limitations and processing breakdown. Ph.d. thesis, Carnegie Mellon University Pittsburgh, PA.

Ginzburg, J. (2012). *The Interactive Stance: Meaning for Conversation*. Oxford: Oxford University Press.

Ginzburg, J., & Cooper, R. (2004). Clarification, ellipsis, and the nature of contextual updates in dialogue. *Linguistics and Philosophy*, *27*(3), 297–365.

Goffman, E. (1981). Forms of Talk. University of Pennsylvania Press.

Gregoromichelaki, E., Kempson, R., Purver, M., Mills, G. J., Cann, R., Meyer-Viol, W., et al. (2011). Incrementality and intention-recognition in utterance processing. *Dialogue and Discourse*, *2*(1), 199–233.

Gregoromichelaki, E., Howes, C., & Kempson, R. (2020). Actionism in syntax and semantics. In *Dialogue and Perception-Extended Papers from DaP2018, CLASP Papers in Computational Linguistics* (Vol. 2, pp. 12–27). Gothenburg: GUPEA.

Healey, P. G. T., & Mills, G. J. (2006). Participation, precedence and co-ordination in dialogue. In *Proceedings of the 28th Annual Conference of the Cognitive Science Society* (pp. 1470–1475).

Healey, P. G. T., Colman, M., & Thirlwell, M. (2005). Analysing multi-modal communication: Repair-based measures of human communicative co-ordination. In J. van Kuppevelt, L. Dybkjaer, & N. Bernsen (Eds.), *Natural, Intelligent and Effective Interaction in Multimodal Dialogue Systems, Text, Speech and Language Technology* (Vol. 30, pp. 113–129). Dordrecht: Kluwer.

Healey, P. G. T., Eshghi, A., Howes, C., & Purver, M. (2011). Making a contribution: Processing clarification requests in dialogue. In *Proceedings of the 21st Annual Meeting of the Society for Text and Discourse, Poitiers*.

Healey, P. G. T., Mills, G. J., Eshghi, A., & Howes, C. (2018). Running repairs: Co-ordinating meaning in dialogue. *Topics in Cognitive Science*, *10*(2), 367–388.

Heldner, M., Hjalmarsson, A., & Edlund, J. (2013). Backchannel relevance spaces. *Nordic Prosody: Proceedings of XIth Conference, Tartu*, *2012*, 137–146.

Hough, J. (2015). Modelling incremental self-repair processing in dialogue. Ph.d thesis, Queen Mary University of London.

Hough, J., Purver, M. (2012). Processing self-repairs in an incremental type-theoretic dialogue system. In *Proceedings of the 16th SemDial Workshop on the Semantics and Pragmatics of Dialogue (SeineDial), Paris, France* (pp. 136–144).

Howes, C. (2012). Coordination in dialogue: Using compound contributions to join a party. Ph.d. thesis, Queen Mary University of London.

Howes, C., Eshghi, A. (2017). Feedback relevance spaces: The organisation of increments in conversation. In *Proceedings of the 12th International Conference on Computational Semantics (IWCS 2017)*. Association for Computational Linguisitics.

Howes, C., Purver, M., Healey, P. G. T., Mills, G. J., & Gregoromichelaki, E. (2011). On incrementality in dialogue: Evidence from compound contributions. *Dialogue and Discourse*, *2*(1), 279–311.

Howes, C., Healey, P. G. T., & Purver, M. (2012a). Whose turn is it anyway? Same- and cross-person compound contributions in dialogue. In *Proceedings of the 34th Annual Meeting of the Cognitive Science Society, Sapporo, Japan* (pp. 1697–1702).

Howes, C., Healey, P. G. T., Purver, M., & Eshghi, A. (2012b). Finishing each other's ... responding to incomplete contributions in dialogue. In *Proceedings of the 34th Annual Meeting of the Cognitive Science Society, Sapporo, Japan* (pp. 479–484).

Kalatzis, D., Eshghi, A., & Lemon, O. (2016). Bootstrapping incremental dialogue systems: Using linguistic knowledge to learn from minimal data. In *Proceedings of the NIPS 2016 Workshop on Learning Methods for Dialogue, Barcelona*.

Kawahara, T., Yamaguchi, T., Inoue, K., Takanashi, K., & Ward, N. (2016). Prediction and generation of backchannel form for attentive listening systems. In *Proceedings of INTERSPEECH* (pp. 2890–2894).

Kempson, R., & Kiaer, J. (2008). Japanese scrambling and the grammar-parser correspondence. In H. Hoshi (Ed.), *Language, Mind and Brain: Perspectives from Linguistics and Cognitive Neuroscience*. Kuroshio: Tokyo.

Kempson, R., Meyer-Viol, W., & Gabbay, D. (2001). *Dynamic Syntax: The Flow of Language Understanding*. Oxford: Blackwell.

Kempson, R., Cann, R., Eshghi, A., Gregoromichelaki, E., & Purver, M. (2015). Ellipsis. In S. Lappin & C. Fox (Eds.), *The Handbook of Contemporary Semantic Theory*. Oxford: Wiley-Blackwell.

Kempson, R., Cann, R., Gregoromichelaki, E., & Chatzikiriakidis, S. (2016). Language as mechanisms for interaction. *Theoretical Linguistics*, *42*(3–4), 203–275.

Kita, S., & Ide, S. (2007). Nodding, aizuchi, and final particles in japanese conversation: How conversation reflects the ideology of communication and social relationships. *Journal of Pragmatics*, *39*(7), 1242–1254.

Kjellmer, G. (2009). Where do we backchannel? On the use of mm, mhm, uh huh and such like. *International Journal of Corpus Linguistics*, *14*(1), 81–112.

Larsson, S. (2015). Formal semantics for perceptual classification. *Journal of Logic and Computation*, *25*(2), 335–369.

Lawler, I., Hahn, F., & Rieser, H. (2017). Gesture meaning needs speech meaning to denote—A case of speech-gesture meaning interaction. *Proceedings of the Workshop on Formal Approaches to the Dynamics of Linguistic Interaction, ESSLLI, 2017*, 42–46.

Mills, G., Healey, P. G. T. (2006). Clarifying spatial descriptions: Local and global effects on semantic co-ordination. In *Proceedings of the 10th Workshop on the Semantics and Pragmatics of Dialogue (SEMDIAL), Potsdam, Germany*.

Mills, G. J. (2007). Semantic co-ordination in dialogue: The role of direct interaction. Ph.d thesis, Queen Mary University of London.

Park, H. W., Gelsomini, M., Lee, J. J., & Breazeal, C. (2017). Telling stories to robots: The effect of backchanneling on a child's storytelling. In *Proceedings of the ACM/IEEE International Conference on Human–Robot Interaction* (pp. 100–108). ACM.

Poppe, R., Truong, K. P., & Heylen, D. (2011). Backchannels: Quantity, type and timing matters. In *International workshop on intelligent virtual agents* (pp. 228–239). Springer.

Purver, M. (2004). The theory and use of clarification requests in dialogue. Ph.D. thesis, University of London.

Purver, M., Cann, R., & Kempson, R. (2006). Grammars as parsers: Meeting the dialogue challenge. *Research on Language and Computation*, *4*(2–3), 289–326.

Purver, M., Howes, C., Gregoromichelaki, E., & Healey, P. G. T. (2009). Split utterances in dialogue: A corpus study. In *Proceedings of the 10th Annual Meeting of the Special Interest Group on Discourse and Dialogue (SIGDIAL), London, UK* (pp. 262–271).

Purver, M., Gregoromichelaki, E., Meyer-Viol, W., & Cann, R. (2010). Splitting the 'I's and crossing the 'you's: context, speech acts and grammar. In *Proceedings of the 14th SemDial Workshop on the Semantics and Pragmatics of Dialogue* (pp. 43–50).

Purver, M., Eshghi, A., & Hough, J. (2011). Incremental semantic construction in a dialogue system. In *Proceedings of the 9th International Conference on Computational Semantics (IWCS), Oxford, UK* (pp. 365–369).

Rühlemann, C. (2007). *Conversation in Context: A Corpus-Driven Approach*. London, New York: Continuum.

Sacks, H., Schegloff, E., & Jefferson, G. (1974). A simplest systematics for the organization of turn-taking for conversation. *Language*, *50*(4), 696–735.

Sato, Y. (2011). Local ambiguity, search strategies and parsing in Dynamic Syntax. In E. Gregoromichelaki, R. Kempson, & C. Howes (Eds.), *The Dynamics of Lexical Interfaces*. Stanford CA: CSLI Publications.

Schegloff, E. (1995). *Parties and talking together: Two ways in which numbers are significant for talk-in-interaction* (pp. 31–42). Situated Order: Studies in the Social Organization of Talk and Embodied Activities.

Schegloff, E., Jefferson, G., & Sacks, H. (1977). The preference for self-correction in the organization of repair in conversation. *Language*, *53*(2), 361–382.

Schegloff, E. A. (1982). Discourse as an interactional achievement: Some uses of 'uh huh' and other things that come between sentences. *Analyzing Discourse: Text and Talk*, *71*, 93.

Tolins, J., & Fox Tree, J. E. (2014). Addressee backchannels steer narrative development. *Journal of Pragmatics*, *70*, 152–164.

Yngve, V. H. (1970). On getting a word in edgewise. In *Papers from the 6th Regional Meeting of the Chicago Linguistic Society* (pp. 567–578).

Yu, Y., Eshghi, A., & Lemon, O. (2016). Training an adaptive dialogue policy for interactive learning of visually grounded word meanings. In *Proceedings of the 17th Annual Meeting of the Special Interest Group on Discourse and Dialogue (SIGDIAL), Los Angeles* (pp. 339–349).