Feedback in Conversation as Incremental Semantic Update[*]

Arash Eshghi
Heriot-Watt University
a.eshghi@hw.ac.uk

Christine Howes
University of Gothenburg
christine.howes@gu.se

Eleni Gregoromichelaki
King's College London
eleni.gregor@kcl.ac.uk

Julian Hough
Bielefeld University
julian.hough@uni-bielefeld.de

Matthew Purver
Queen Mary University of London
mpurver@qmul.ac.uk

Abstract

In conversation, interlocutors routinely indicate whether something said or done has been processed and integrated. Such feedback includes *backchannels* such as 'okay' or 'mhm', the production of a next relevant turn, and repair initiation via *clarification requests*. Importantly, such feedback can be produced not only at sentence/turn boundaries, but also sub-sententially. In this paper, we extend an existing model of incremental semantic processing in dialogue, based around the Dynamic Syntax (DS) grammar framework, to provide a low-level, integrated account of backchannels, clarification requests and their responses; demonstrating that they can be accounted for as part of the core semantic structure-building mechanisms of the grammar, rather than via higher level pragmatic phenomena such as intention recognition, or treatment as an "unofficial" part of the conversation. The end result is an incremental model in which words, not turns, are seen as procedures for contextual update and backchannels serve to align participant semantic processing contexts and thus ease the production and interpretation of subsequent conversational actions. We also show how clarification requests and their following responses and repair can be modelled within the same DS framework, wherein the divergence and re-alignment effort in participants' semantic processing drives conversations forward.

1 Introduction

In conversation, interlocutors provide frequent feedback about whether something said can be taken as understood. For Clark (1996), a crucial point of advancing the joint project of dialogue is establishing that we are sufficiently coordinated to continue, a process called "grounding", which uses, for example, backchannel responses (such as 'mm', example 1:2127, or 'yeah') and non-linguistic cues (e.g. nods and smiles).[1] Alternative responses indicate processing difficulties or lack of coordination and signal a need for clarification or repair (example 1:2112).

Clark and Schaefer (1989) present a model of *contributions* in dialogue that consist of both a *presentation* and an *acceptance* phase. In the acceptance phase, listeners can display evidence of understanding at various levels, from continued attention to verbatim repetition, with backchannels being one possible type (see also Allwood et al., 1992). The acceptance phase can also be used to clear up sources of misunderstanding, as with clarification requests. Traum (1994) develops a formal model of grounding in which *grounding acts* at the level of individual utterances build up *discourse units*, at which level core speech acts are realised through being grounded. More recently, Ginzburg (2012) provides a substantial dialogue model based on embedding the grammar under an utterance processing protocol, modelling updates of interlocutors' information states. Grounding or clarification processes rely on a notion of

[1]We focus on verbal feedback here; however, we believe that the analysis also applies to non-verbal feedback e.g. nodding.

"locutionary proposition", a linguistic sign, specified with an appropriate illocutionary force through the grammar. Such propositions become the elements manipulated during the grounding process, resulting in either acceptance in the common ground or the generation of clarification. Elements providing feedback or repair are assigned lexical entries or construction types that presuppose the derivation of such propositions in the context.

However, importantly, such feedback can be produced sub-sententially and during an ongoing turn (as with the backchannels in (1) lines 2127 and 2129), illustrating the crucially incremental nature of coordination in human interaction, without the need to invoke propositional contents at all stages.

(1)   Example dialogue from the British National Corpus (BNC) KPU[2]

Gearoid	2111	We went to see something called the Wedding Banquet.
Anne-Marie	2112	Called the Wedding [Banquet]?
Gearoid	2113	[Banquet].
Anne-Marie	2114	Really?
⋮	⋮	⋮
Gearoid	2126	It's about these two chaps living in New York ⟨pause⟩ one is American and the other is er ⟨pause⟩ Hong Kong, no Chinese ⟨pause⟩ and they're an affair, they're [living]
Anne-Marie	2127	[Mm].
Gearoid	2128	together, very [sort of]
Anne-Marie	2129	[Mm].
Gearoid	2130	you know, kind, one is an architect and the other ⟨pause⟩ the Chinese guy is a property dealer.

Moreover, in most dialogue models, grounding or clarification generation introduce a speaker/hearer asymmetry in their representations in that speakers are assumed to be omniscient regarding their own utterances (which seems to preclude backchannels/clarifications addressed to self), as a result of complete, clear plans/intentions guiding their production (e.g. Poesio and Rieser, 2010). However, the joint nature of dialogue actions becomes evident in the fact that, besides listenership, turn-managing, understanding, and acceptance backchannel feedback and its elicitation, like any mechanism in interaction, have perlocutionary effects in conversation, post-hoc characterisable both as "intended" and "unintended" (2):

	A:	John...uh... yes, John, yeah?
	B:	mhm, mhm
	A:	he went to the party yest....
(2)	B:	yeah, yes, ok.
	A:	He saw your sister with.
	B:	yes, yes.
	A:	Will you stop hurrying me along? I'll tell the story at my own pace.

Similarly, repetitions of phrases/sentences in dialogue serve various feedback functions (both for self and other) besides clarification. These range from surprise or disbelief indicators to delaying functions, and, like backchannels, their content cannot always be explicated as a full propositional intention without some loss of the impression or effect shared in context (example 3).

	A:	1	Nirma was at the party.
	A:	2a	Nirma! / Nirma? / Who? (Why Nirma? / Sorry, I meant Irma.)
	B:	2b	Nirma eh? I knew it.
(3)	B:	2c	Nirma! Oh how nice.
	B:	2d	Nirma... Wait, I know this name.
	B:	2e	Nirma, Nirma, I see.

---

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

Models of grounding have been recently explored in practical dialogue systems. However, these systems either explore the positioning of backchannels based on low-level features (e.g. Cathcart et al., 2003; Gravano and Hirschberg, 2009); or rely on a notion of feedback that incorporates reasoning about the intentions, mental states or goals of one's interlocutor (e.g. Visser et al., 2014; Buschmeier and Kopp, 2013; Wang et al., 2011). The former type of system may allow a dialogue model to sound 'more human', but do not give any insight into why feedback occurs where it does; and, in the latter, full intention recognition requires a level of complexity that corpus studies on repair (Purver et al., 2003; Colman and Healey, 2011), backchannels (Kjellmer, 2009), and conversational analysis of multiparty dialogue (Goodwin, 1979; Koschmann and LeBaron, 2003) suggest are unnecessary as a prerequisite in natural conversation. In contrast, under the view pioneered by Ginzburg (2012) and Poesio and Rieser (2010), all such phenomena require an account that integrates their linguistic features with their functions in dialogue because of various form-content constraints. However, even if this is the general perspective assumed here, examples (1)-(3) show that the grammar needs to be able to provide an account of such feedback and repair mechanisms that remains flexible enough to be put to various situated uses without requiring specific propositional contents to be derived for single words or phrases. This is a significant requirement given that both the content (Clark, 1996) and the scope of a feedback contribution is highly underspecified (as also pointed out by an anonymous IWCS reviewer), rather than determinable through use of fixed constructions. Moreover, backchannels and clarification requests, like anaphora and ellipsis antecedents, can be provided through actions employing various modalities, like eye-gaze, posture, head movement and bodily gestures (Goodwin, 1986). For this reason the grammar model needs to be able to integrate input from various sources without rules that confine such input to linguistic signs (cf Ginzburg (2012) whose locutionary propositions preclude such unification).

In this paper we first outline the formal model of Dynamic Syntax with Type Theory with Records (DS-TTR); we then present an extension to the model and show how it can provide a low-level, semantic model of feedback phenomena, in particular backchannels and clarification interaction, without recourse to dialogue moves/acts or reasoning about intentions. The model is illustrated using variations on example (4).

(4)   Example dialogue from BNC KPY
    A   1006   Er, the doctor
    B   1007   Chorlton?
    A   1008   Chorlton, mhm, he examined me, erm, he, he said now they were on about a slide ⟨unclear⟩ on my heart.

2   Dynamic Syntax and Type Theory with Records (DS-TTR)

Dynamic Syntax is an action-based grammar formalism, which models the word-by-word incremental processing of linguistic input. Unlike many other formalisms, DS models the incremental linear construction of *interpretations* without recognising an independent level of syntactic representation. Thus, the output for any given string of words is a purely *semantic* tree representing its predicate-argument structure; tree nodes correspond to terms in the lambda calculus, decorated with labels expressing their semantic type (e.g. $Ty(e)$) and logical formulae as record types of the Type Theory with Records framework (TTR, see below); beta-reduction determines the type and formula at a mother node from those at its daughters (Figure 1).

These trees can be *partial*, containing unsatisfied *requirements* potentially for any element, for example, node labels (e.g. $?Ty(e)$, a requirement for future development to $Ty(e)$), and contain a *pointer*, ◇, labelling the node currently under development. Grammaticality is defined as processability 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.[3]

---

[3]In this paper, we exclude all considerations of tense/aspect for simplicity. But see Cann (2011).

2.1 Actions in DS

The parsing process is defined in terms of conditional *actions*: procedural specifications for monotonic tree/string development. These can be either general structure-building principles (*computational actions*) or language-specific actions induced by parsing particular lexical items (*lexical actions*).

Computational actions These form a small, fixed set of IF..THEN..ELSE action specifications. Some merely encode the properties of the lambda calculus and the logical tree formalism (LoFT, Blackburn and Meyer-Viol, 1994): e.g. THINNING (removal of satisfied requirements), COMPLETION (moving the pointer up and out of a lower sub-tree once all requirements therein are satisfied, see Figure 1), and ELIMINATION (beta-reduction of daughter nodes at the mother). Others reflect the fundamental predictivity and dynamics of DS: *ADJUNCTION introduces a single unfixed node with underspecified tree position (replacing feature-passing concepts for phenomena like long-distance dependency); LINK-ADJUNCTION builds a paired ("linked") tree corresponding to semantic conjunction (licensing relative clauses, apposition and more). These actions represent possible processing strategies, applying optionally at any stage of a parse if their preconditions are met.
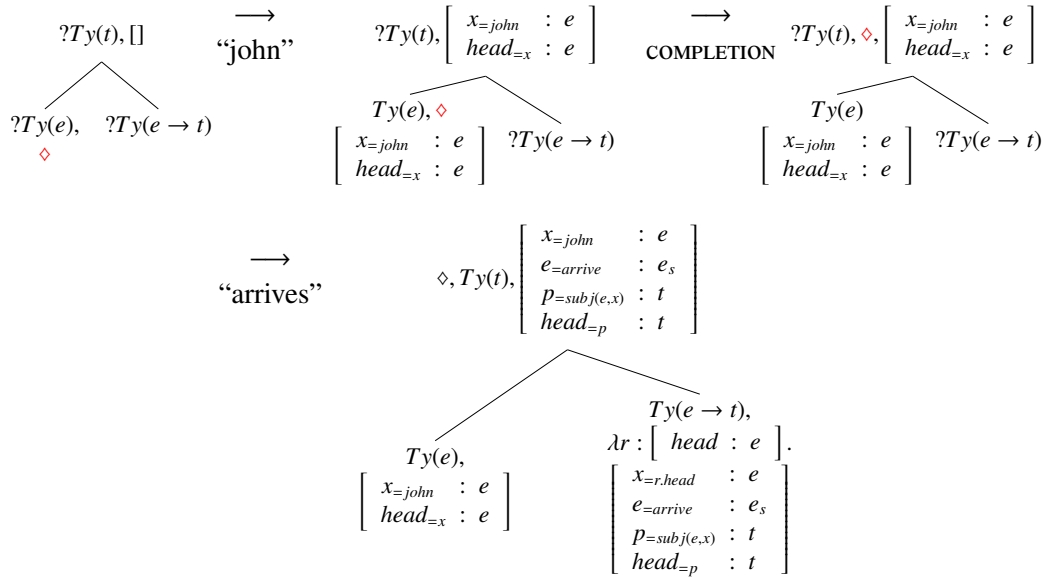
Figure 1: Incremental parsing in DS-TTR: *"John arrives"*

Lexical actions The lexicon associates word forms with lexical actions; like computational actions, these are also sequences of tree-update actions in an IF..THEN..ELSE format, composed of *atomic* tree-building actions such as `make`, `go`, `put`. `make` creates a new daughter node, `go` moves the pointer, and `put` decorates the pointed node with a label. Figure 2 shows an example for a proper noun, *John*. The action checks whether the pointed node (marked as $\diamond$) has a requirement for type e; if so, it decorates it with type e (thus satisfying the requirement), formula $John'$ and the bottom restriction $\langle\downarrow\rangle\bot$ (meaning that the node cannot have any daughters). Otherwise (if no requirement $?Ty(e)$), the action aborts, meaning that the word *'John'* cannot be parsed in the context of the current tree.

Figure 2: Lexical action for the word 'John'

Figure 3: DS parsing as a graph: actions (edges) are transitions between partial trees (nodes).

2.2 Type Theory with Records

Type Theory with Records (TTR) is an extension of standard type theory shown useful in semantics and dialogue modelling (Cooper, 2005; Ginzburg, 2012). For us here, it provides the logical formalism in which meanings are expressed (Purver et al., 2011; Hough and Purver, 2012; Eshghi et al., 2012). Given its fine-grained, structured representations (see below), it has also been used to encode the linguistic, 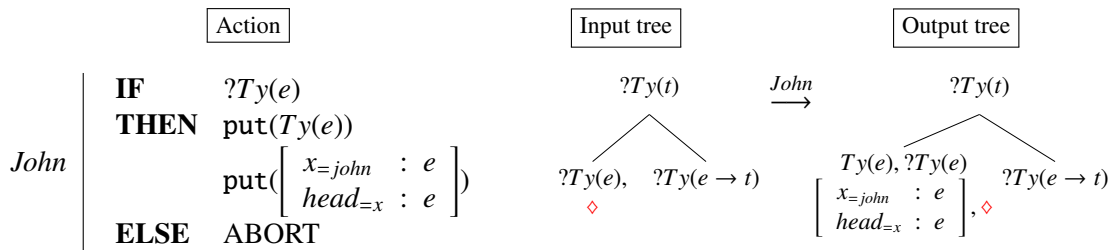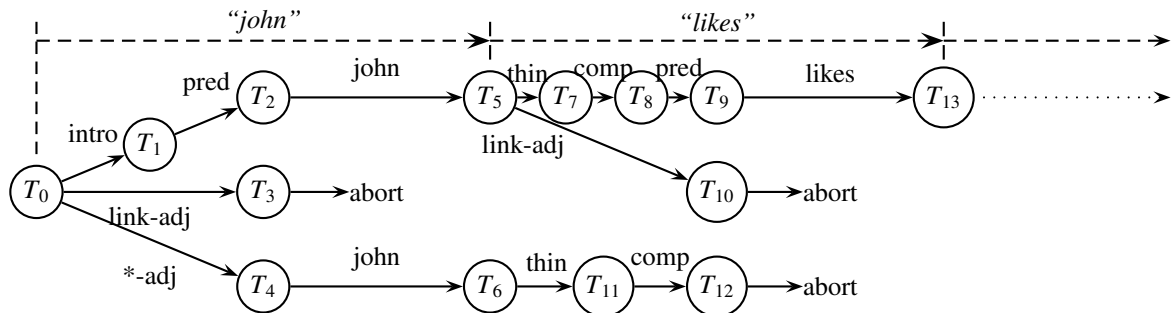and non-linguistic context of an utterance (Purver et al., 2010; Dobnik et al., 2012). This is important for us here since such an integration provides for non-linguistic elicitations and provisions of feedback.

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

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 using the syntactic sugar $[\, l_{=a} : T \,]$. Fields can also be *dependent* on fields preceding them (i.e. higher) in the record type – e.g. in $\begin{bmatrix} l_1 & : & T_1 \\ l_{2=a} & : & T_2 \\ l_{3=p(l_2)} & : & T_3 \end{bmatrix}$. Importantly for us here, 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 $T_1 \sqsubseteq T_2$.

Following Purver et al. (2011), we assume that DS tree nodes are decorated with RTs, and corresponding lambda abstracts representing functions from RT to RT (e.g. $\lambda r : [\, l_1 : T_1 \,].[\, l_{2=r.l_1} : T_1 \,]$ where $r.l_1$ is a *path* expression referring to the label l_1 in r) – see Figure 1. TTR's subtype relation allows a record type at the root node to be inferred for any partial tree, and incrementally further specified via subtyping as parsing proceeds (Hough and Purver, 2012).

2.3 Graph-based Parsing & Generation

In parsing, given a sequence of words $(w_1, w_2, ..., w_n)$, the parser starts from the *axiom* tree T_0 (a requirement to construct a complete propositional tree, $?Ty(t)$), and applies the corresponding lexical actions $(a_1, a_2, \ldots, a_n)$, optionally interspersing computational actions. This can be modelled as a directed acyclic graph (DAG) rooted at T_0, with partial trees as nodes, and computational and lexical actions as edges (i.e. transitions between trees) (Sato, 2011). Figure 3 shows an example: here, *intro*, *pred* and **adj* correspond to the computational actions INTRODUCTION, PREDICTION and *-ADJUNCTION respectively; and 'john' is a lexical action. Different DAG paths represent different parsing strategies, which may succeed or fail depending on how the utterance is continued. Here, the path $T_0 - T_5$ will succeed if 'John' is the subject of an upcoming verb ("John upset Mary"); $T_0 - T_6$ will succeed if 'John' turns out to be a left-dislocated object ("John, Mary upset").

This incrementally constructed DAG makes up the entire *parse state* at any point. The rightmost nodes (i.e. partial trees) make up the current maximal semantic information; these nodes with their paths back to the root (tree-transition actions) make up the *linguistic context* for ellipsis and pronominal construal (Purver et al., 2011).

This fine-grained DAG can also be seen as subsumed by a coarser-grained DAG at the word level; at this level, edges represent words, with nodes representing sets of (partial) trees, which are the right-most

nodes in the more fine-grained parse state DAG after processing that word. For Figure 3, this level would consist of two edges: one for *"john"* connecting a node $\{T_0\}$ to a node $\{T_5, T_6\}$; and one for *"likes"* connecting that node $\{T_5, T_6\}$ to a node $\{T_{13}\}$. This higher-level representation obscures grammatical parsing details, but is more compatible with speech recogniser input and dialogue manager output in a practical system. In our explanations below, we will use this coarser-grained representation.

3 A semantic model of feedback

The extension to DS-TTR we make here is to represent the state of multiple dialogue participants as they jointly construct and negotiate meaning in dialogue by providing different forms of feedback. To model grounding states, we make use of the parser/generation context DAG (see Figure 3) for a given dialogue participant, augmented with two coordination pointers (different to the tree pointer mentioned above): one pointer, dubbed the *self-pointer*, ♦, indicates the node in the DAG which that dialogue participant has provided evidence for reaching (by producing any contributing output, including backchannelling, answering a question, extending another participant's utterance, or repair initiation). The operation of this pointer can also be regarded as a self-monitoring device, which can trigger self-addressed backchannels and repairs. The second coordination pointer, which we term the *other-pointer*, ◊, indicates the node in the DAG which one's interlocutor(s) have provided evidence for reaching.

This allows us to model feedback in the form of backchannels, CRs, continuations and answers to questions, or indeed any local use of context-dependency. Table 1 shows some of the forms this feedback might take. The model is defined within the incremental process of joint semantic construction, without recourse to higher-level pragmatic reasoning, dialogue acts, or intention recognition.

On this account, if the two coordination pointers are on the same DAG node, any (sub-)utterance on the DAG path from this doubly pointed node back to the root can be taken to be grounded. More generally, the intersection of the ♦ → root path and ◊ → root path is grounded; the remaining ◊ → root path is as yet ungrounded; and other paths are *repaired* (see below). Divergence of the two pointers represents the source for forward momentum in dialogue: something must be done by either of the parties in order to align these pointers. This is conceptually analogous to Matheson et al. (2000)'s notion of *obligation* or Ginzburg (2012)'s *discursive potential*, but operates without recourse to dialogue acts such as acceptance, or rejection. As will become clear below, linguistic signs of rejection here do not necessarily correspond to denying, or discussing a proposition, but instead indicate abandonment of a DAG branch. Under this general view, linguistic elements can unproblematically operate sub-propositionally, for instance, establishing a referent before the proposition involving that referent is complete.

	Local			**Non-Local**		
Confirmed	(a)	A: B: A:	The doctor Chorlton? Chorlton, mhm, he examined me	(b)	A: B: A:	The doctor examined me Chorlton? Chorlton, mhm, he examined me
Repaired	(c)	A: B: A: B:	The doctor Chorlton? no, Fitzgerald uh-huh	(d)	A: B: A: B:	The doctor examined me Chorlton? no, Fitzgerald uh-huh

Table 1: Local & Non-Local Clarification Interaction

3.1 Backchannels

Backchannels as DAG pointer movement As shown in Figure 4, producing any utterance has the effect of moving one's self-pointer ♦ along the DAG; parsing something someone else says moves one's other-pointer ◊. As one person speaks and the other listens, the two pointers diverge in position; the intervening material is ungrounded (see Poesio and Traum (1997)'s *ungrounded discourse units*). An analysis of backchannels as signs associated with null parsing actions (actions which contribute no new

Utterance	Context-final Semantics	A's Context After Utterance

(a) A: The doctor
$$\left[\begin{array}{l} r \quad : \left[\begin{array}{ll} x & : e \\ p_{=doctor(x)} & : t \end{array}\right] \\ x_{=\iota(r.x,r)} \quad : e \end{array}\right]$$

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

(b) B: mhm
$$\left[\begin{array}{l} r \quad : \left[\begin{array}{ll} x & : e \\ p_{=doctor(x)} & : t \end{array}\right] \\ x_{=\iota(r.x,r)} \quad : e \end{array}\right]$$

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

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

The   *doctor* $\diamond$ *he*   *examined me* $\blacklozenge$
$s_0 \rightarrow s_1 \rightarrow s_2 \rightarrow s_3 \rightarrow s_4 \rightarrow s_5$
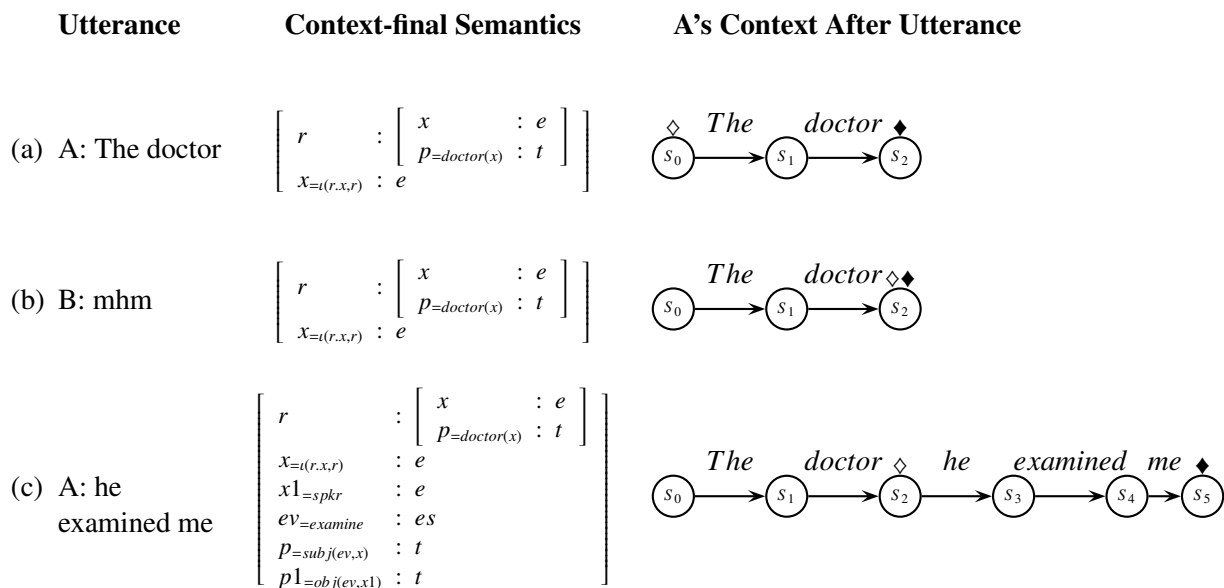
Figure 4: Backchannels as movement of context DAG coordination pointers. From A's perspective.

semantic information to the partial DS tree under construction) would therefore provide a minimal account: parsing such a trivial action produced by an interlocutor after producing some output will move the other-pointer to catch up with the self-pointer; generating such an action after hearing an interlocutor's input will move the self-pointer to catch up with the other-pointer (see Figure 4(a-b)). Backchannels can therefore be seen as minimal utterances to carry out this pointer convergence; they act as grounding signals by ensuring that both pointers are at the same node in the DAG.

Backchannels as semantic compilation Such a simple account, however, fails to capture the distribution of backchannels: it would predict that they should be equally likely at any point, and this is not the case. A richer account is available if we also take backchannels to be associated with non-trivial lexical entries which apply the standard DS computational action of COMPLETION - the movement of the DS tree pointer from a daughter node with no outstanding semantic requirements to its mother (see Fig. 1). This action is taken to be freely available in DS, as a necessary part of the basic tree compilation process (see Kempson et al., 2001); however, additionally associating it with particular lexical entries will ensure that it is applied if it can be, whenever such a word is parsed or generated.

Under this account, parsing or generating a backchannel causes the COMPLETION action to be run; backchannels are therefore useful at points in the parse where a sub-tree (semantic constituent) has just been completed. Not only does this have the coordinating effect of aligning DAG pointers (as above), it also reduces parse-state ambiguity, reducing the partial-tree set currently under consideration to the subset where COMPLETION has been carried out (if such a subset exists). For example, in the dialogue, "A: Mary, my friend B: mhm A: . . . ", B's backchannel will cause A to eliminate DAG paths (or make them less likely to be followed) that are not compatible with COMPLETION, thus hindering, e.g. further qualifications of 'Mary', such as a relative clause (because the DS tree pointer has moved out of the 'Mary' sub-tree). Although the existence of such preferences is an empirical question, it seems intuitive because B has presumably resolved the 'Mary' reference successfully.

Backchannel-relevance spaces (Heldner et al., 2013) are therefore predicted to be places where COMPLETION is possible, without having to postulate this as a separate pragmatic rule, or see them as associated with a specified dialogue update function. It is also predicted that use of such processing strategies carries risks, in that they can be misconstrued, misused or abused. For example, if a backchannel is produced where, from the perspective of the speaker, COMPLETION should not be available (e.g. mid-NP constituent), this may indicate that a participant's DAG is not coordinated with their interlocutor's. In various sociolinguistic studies, uncoordinated feedback is taken to indicate lack of attention, rapport or interest. On the other hand, this flexibility can also be exploited for various effects. Kjellmer (2009) discusses in-
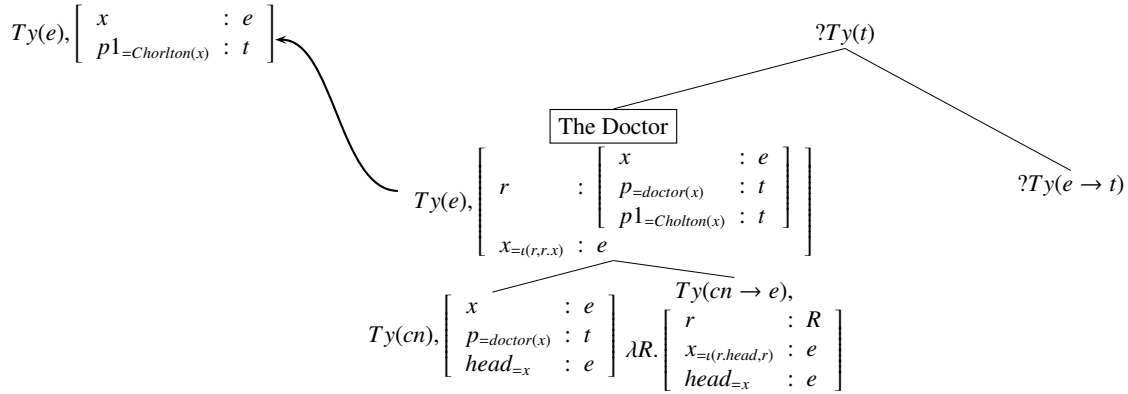
$$Ty(e), \begin{bmatrix} x & : e \\ p1_{=Chorlton(x)} & : t \end{bmatrix} \qquad\qquad ?Ty(t)$$

The Doctor

$$Ty(e), \begin{bmatrix} r & : \begin{bmatrix} x & : e \\ p_{=doctor(x)} & : t \\ p1_{=Cholton(x)} & : t \end{bmatrix} \\ x_{=\iota(r.r.x)} & : e \end{bmatrix} \qquad ?Ty(e \to t)$$

$$Ty(cn), \begin{bmatrix} x & : e \\ p_{=doctor(x)} & : t \\ head_{=x} & : e \end{bmatrix} \qquad \lambda R. \begin{bmatrix} Ty(cn \to e), \\ r & : R \\ x_{=\iota(r.head,r)} & : e \\ head_{=x} & : e \end{bmatrix}$$

Figure 5: Processing *Chorlton?* in "A: *the doctor* B: *Chorlton?*"

terruptive feedback cases where the next element (after which COMPLETION would be available) is highly predictable. In our DS modelling this would indicate that the listener, through the grammar's generation of predictions, is able to achieve conceptual access and has reached the point at which COMPLETION can apply without all the words being articulated (analogously to Howes et al., 2012, which showed that people responded to incomplete utterances as if what was in the surface had been completed if the 'missing' material was highly predictable). This might give the surface appearance of an abandoned but, nevertheless, grounded utterance.[4] These possibilities highlight important points. Firstly, due to the situatedness of interaction, feedback has an open-ended range of functions resulting in potential participant mismatches. Consequently, a "feedback" mechanism may be used as a tool to perform another (implicit) dialogue act that moves the conversation forward or achieves particular perlocutionary effects (see, e.g., (2) earlier). Finally, from a modelling point of view, these effects emphasise that, not only incremental parsing, but also predictive processing, when incorporated as part of the grammar, can provide for a wider range of effects and generalisations across phenomena.

3.2 Clarification Interaction

Processing Clarification Requests In DS-TTR, elliptical clarification requests (CR), taking the form of NP-repetitions, or appositions that aim to clarify an NP's content, are processed as *extending* a semantic tree in context. This is done using the LINK-ADJUNCTION computational action (see Section 2) motivated independently for processing relative clauses and adjuncts (Cann et al., 2005). This leads to the creation of a paired LINKed tree of root node type e ($Ty(e)$), which is linked to the $Ty(e)$ node annotated with the content of the NP being clarified (see Fig. 5). Once the LINKed tree is complete, its content is combined through record type *extension* (TTR variant of conjunction, see Section 2.2) with the content of the NP being clarified. NP content is modelled in DS-TTR as the compilation of terms of the epsilon calculus. This provides the potential of indefinitely extending the restrictors of such terms, which is exploited here for accumulating contents (sometimes trivially) resulting from clarification sequences.

CRs can appear both locally, as in Table 1, (a,c), or non-locally from their antecedent, as in (b,d). Nevertheless, we model both in the same way, as extensions of the relevant NP's content. As Fig. 5 shows, processing *Chorlton?* locally involves building a LINKed tree from its antecedent, "the doctor". Once this LINKed tree is complete, the Record Type (RT) representing the content of *the doctor* is extended with that of "Chorlton", leading to an annotation on that node marked as The Doctor (which designates a unique individual who is a doctor and whose name is Chorlton). Parsing/production can then continue as normal, thus setting up the context in which the CR response is processed (whether that response is positive as in Table 1 (a,b) or negative as in (c,d)).

Non-local CRs require backtracking along the current DAG branch to a point where the CR can be parsed in the same way (see Fig. 6 and for further details Hough and Purver's (2012) model of self-repairs). This backtracking is triggered by the inability to parse/produce in the local context (or a very

---

[4]Thanks to an anonymous IWCS reviewer for pointing out this phenomenon, which, in our view, parallels our DS modelling of compound contributions in Purver et al. (2010).

low probability path having to be taken). The actions backtracked over (the DAG edges traversed) are then re-applied once the CR has been parsed or produced (i.e. when suitably indicated by punctuation or intonation). This "action replay", independently motivated for other forms of ellipsis (Kempson et al., 2014; Gargett et al., 2009), leads to the full content of the CR being established as before.

Contextual divergence   The above account of how CRs are parsed and produced does not sufficiently capture all aspects of a clarificational exchange. What is also needed is an account of negative or positive responses to CRs, and the modelling of speaker/hearer context divergence during a clarification sequence. Having introduced the self- and other-coordination DAG pointers motivated earlier for an account of backchannels, we can now provide an incremental model of speaker/hearer interaction during a clarificational exchange in terms of how these pointer positions initially diverge to process a CR, and then converge again, thus re-establishing grounding, once the CR and its response have been integrated.

Figure 6 shows the incremental updates arising in the clarifier B's context in example (d) in Table 1, a case of a non-local CR which requires backtracking.[5] Initially, B successfully parses A's utterance, thus moving the other-pointer to the right-most node of his DAG. Not having secured a referent for "the doctor" with enough certainty, he then aims to produce the CR, *Chorlton?*, which involves backtracking to "the doctor" node in order to produce it. At this juncture A and B's contexts have diverged: A's self-pointer appears at the rightmost DAG edge while B has not grounded that edge. B's production of the CR causes A to have to parse it. This serves to re-align pointer positions for A and B, the result of which is both of them focussing on "the doctor"-sub-tree as the source of the misalignment. A can now offer a confirmatory or a negative response to the CR. (c), in Fig. 6 illustrates the latter case, with the utterance of *no* reflecting the abandonment of the "Chorlton?" branch, rather than the denial or rejection of a propositional content. This is followed by a correction of B's CR, thus forcing B's white other-pointer out of the "Chorlton?" branch, and inducing the construction of the new, "Fitzgerald" branch. At this juncture, B's self- and other-pointers are on different branches. This can be taken as representing the requirement for further action to be taken in order to realign pointer positions. Especially for B, whose pointer is now on an abandoned repaired branch, this can constitute an obligation to ground the new information provided by A's repair *Fitzgerald*, thus accounting for the forward momentum created by the negative response. Note how the repaired branch is still part of the context, as like Ginzburg et al. (2014) we take the repaired material as grounded. B's final backchannel, in 6(d), then serves to realign his two pointers, signalling acceptance to A, who, having processed the backchannel moves her white, other-pointer to the same node, $s10$, thus ending the clarification sequence with the achievement of a realignment of A's and B's contexts.

An alternative parsing path is illustrated in (e) of Figure 6. It represents the case where A, after the clarification in 6(b), confirms that the doctor is in fact Chorlton. This simply involves, for B, moving his other-pointer to the end of the "Chorlton?" branch, thus confirming the referent of *the doctor* as Chorlton. This, unlike the negative response in 6(c) which necessitated rejecting already established branches and pointer divergence, ends the clarification sequence. Both alternatives end up with A's and B's contexts aligned as the result of repair and backchannelling and set for the continuation of the dialogue.

4   Conclusions

We have presented a word-by-word incremental account of the coordination phenomena backchannels and clarification interaction within a semantic parsing framework. Given a psychologically realistic time-linear processing model must maintain multiple live options in parallel, such phenomena can provide substantial gains for a coupled parser/generator system by ensuring localisation of possible repair points within recent, ungrounded material. However, such gains can only be modelled if the system supports fully incremental processing, without necessary recourse to sentential or propositional constructs or pre-specified illocutionary forces. Such an approach explains not only the pervasiveness of coordination strategies, but their availability for flexible, creative use by interlocutors at the pragmatic level to express and elicit various social and affective effects.

---

[5]For space reasons we do not here include the clarification recipient A's point of view.

Utterance	Context-final Semantics	B's Context After Utterance

(a) A: The doctor 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* *examined* *me* ◇
$s_0 \rightarrow s_1 \rightarrow s_2 \rightarrow s_3 \rightarrow s_4$

(b) B: Chorlton?

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

The doctor examined me ◇
$s_0 \rightarrow s_1 \rightarrow s_2 \dashrightarrow s_3 \dashrightarrow s_4$
Chorlton?
$s_5 \rightarrow s_6 \rightarrow s_7$ ♦
examined me

(c) A: (no,) Fitzgerald

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

The doctor examined me
$s_0 \rightarrow s_1 \rightarrow s_2 \dashrightarrow s_3 \dashrightarrow s_4$
Chorlton?
Fitzgerald $s_5 \dashrightarrow s_6 \dashrightarrow s_7$ ♦
examined me
$s_8 \rightarrow s_9 \rightarrow s_{10}$ ◇
examined me

(d) B: uh-huh

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

The doctor examined me
$s_0 \rightarrow s_1 \rightarrow s_2 \dashrightarrow s_3 \dashrightarrow s_4$
Chorlton?
Fitzgerald $s_5 \dashrightarrow s_6 \dashrightarrow s_7$
examined me
$s_8 \rightarrow s_9 \rightarrow s_{10}$ ♦◇
examined me

(e) A (accept): yes/mhm

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

The doctor examined me
$s_0 \rightarrow s_1 \rightarrow s_2 \dashrightarrow s_3 \dashrightarrow s_4$
Chorlton?
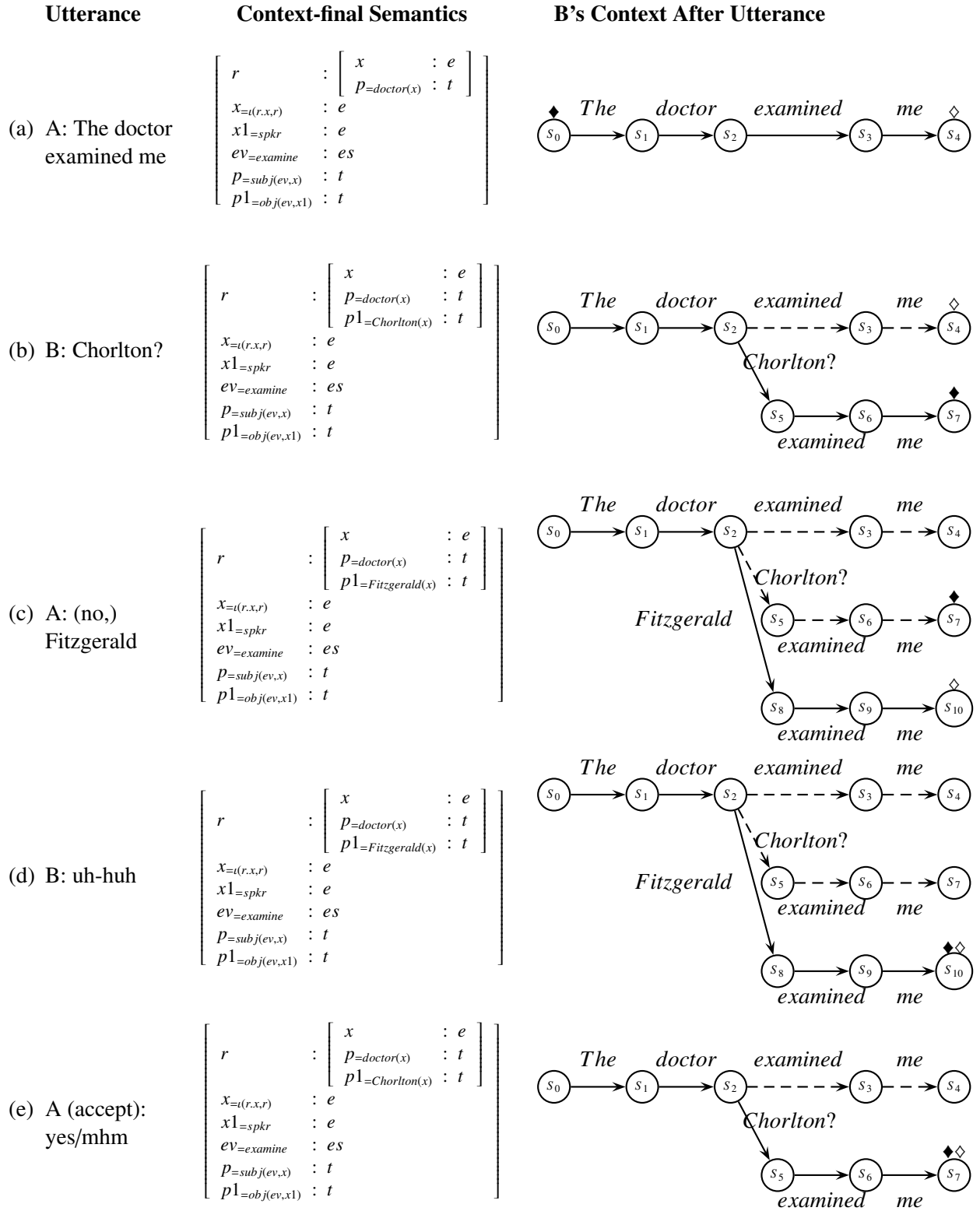$s_5 \rightarrow s_6 \rightarrow s_7$ ♦◇
examined me

Figure 6: Clarification Interaction in DS-TTR from B's perspective. Turn (e) comes after (b)

References

Allwood, J., J. Nivre, and E. Ahlsén (1992). On the semantics and pragmatics of linguistic feedback. *Journal of Semantics 9*(1), 1–26. Also published as Gothenburg Papers in Theoretical Linguistics 64.

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

Buschmeier, H. and S. Kopp (2013). Co-constructing grounded symbols–feedback and incremental adaptation in human-agent dialogue. *KI-Künstliche Intelligenz 27*(2), 137–143.

Cann, R. (2011). Towards an account of the English auxiliary system: building interpretations incrementally. In R. Kempson, E. Gregoromichelaki, and C. Howes (Eds.), *Dynamics of Lexical Interfaces*. Chicago: CSLI Press.

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

Cathcart, N., J. Carletta, and E. Klein (2003). A shallow model of backchannel continuers in spoken dialogue. In *Proceedings of the 10th conference of the European Chapter of the Association for Computational Linguistics (EACL)*, pp. 51–58.

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

Clark, H. H. and E. F. Schaefer (1989). Contributing to discourse. *Cognitive Science 13*(2), 259–294.

Colman, M. and P. Healey (2011, July). The distribution of repair in dialogue. In C. Hoelscher and T. Shipley (Eds.), *Proceedings of the 33rd Annual Conference of the Cognitive Science Society*, Boston, MA, pp. 1563–1568.

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

Dobnik, S., R. Cooper, and S. Larsson (2012). Modelling language, action, and perception in type theory with records. In *Proceedings of the 7th International Workshop on Constraint Solving and Language Processing (CSLP-12)*, pp. 51–63.

Eshghi, A., J. Hough, M. Purver, R. Kempson, and E. Gregoromichelaki (2012). Conversational interactions: Capturing dialogue dynamics. In *From Quantification to Conversation*, Volume 19 of *Tributes*, pp. 325–349. College Publications.

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

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

Ginzburg, J., R. Fernández, and D. Schlangen (2014, June). Disfluencies as intra-utterance dialogue moves. *Semantics and Pragmatics 7*(9), 1–64.

Goodwin, C. (1979). The interactive construction of a sentence in natural conversation. In G. Psathas (Ed.), *Everyday Language: Studies in Ethnomethodology*, pp. 97–121. New York: Irvington Publishers.

Goodwin, C. (1986). Audience diversity, participation and interpretation. *Tex 6*(3), 283–316.

Gravano, A. and J. Hirschberg (2009). Backchannel-inviting cues in task-oriented dialogue. In *INTERSPEECH*, pp. 1019–1022.

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

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

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

Kempson, R., R. Cann, A. Eshghi, E. Gregoromichelaki, and M. Purver (2014). Ellipsis. In S. Lappin and C. Fox (Eds.), *Handbook of Contemporary Semantic Theory*. Willey-Blackwell Publications.

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

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.

Koschmann, T. and C. D. LeBaron (2003). Reconsidering common ground: Examining Clark's contribution theory in the OR. In *Proceedings of the 8th European Conference on Computer Supported Cooperative Work*, pp. 81–98. Kluwer.

Matheson, C., M. Poesio, and D. Traum (2000, April). Modeling grounding and discourse obligations using update rules. In *Proc. 1st Meeting of the North American Chapter of the Association for Computational Linguistics (NAACL)*, Seattle.

Poesio, M. and H. Rieser (2010). Completions, coordination, and alignment in dialogue. *Dialogue and Discourse 1*, 1–89.

Poesio, M. and D. Traum (1997). Conversational actions and discourse situations. *Computational Intelligence 13*(3), 309–347.

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

Purver, M., J. Ginzburg, and P. G. T. Healey (2003). On the means for clarification in dialogue. In R. Smith and J. van Kuppevelt (Eds.), *Current and New Directions in Discourse & Dialogue*, pp. 235–255. Kluwer Academic Publishers.

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

Sato, Y. (2011). Local ambiguity, search strategies and parsing in Dynamic Syntax. In E. Gregoromichelaki, R. Kempson, and C. Howes (Eds.), *The Dynamics of Lexical Interfaces*, pp. 205–233. CSLI.

Traum, D. (1994). *A Computational Theory of Grounding in Natural Language Conversation*. Ph. D. thesis, U. Rochester.

Visser, T., D. Traum, D. DeVault, and R. op den Akker (2014). A model for incremental grounding in spoken dialogue systems. *Journal on Multimodal User Interfaces 8*(1), 61–73.

Wang, Z., J. Lee, and S. Marsella (2011). Towards more comprehensive listening behavior: beyond the bobble head. In *Intelligent Virtual Agents*, pp. 216–227. Springer.