Feedback relevance spaces: The organisation of increments in conversation

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

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

Abstract

Feedback such as backchannels and clarification requests can occur 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 provide a low-level, semantic processing model of where feedback ought to be licensed. The model can account for cases where feedback occurs at FRSs, and 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.

1 Introduction

Dialogue is co-constructed by multiple interlocutors with the traditional split between *speaker* and *hearer* inadequate to describe how this proceeds. Even in monological contexts (e.g. lectures), 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 (e.g. 'mm', as in (1):5144,[1] or 'yeah' (1):5146) including non-linguistic cues (e.g. nods).[2] Other responses indicate processing difficulties or lack of coordination and signal a need for repair ((1):5158) (Bavelas et al., 2012). Further, feedback affects how the conversation unfolds even when it does not contribute any semantic content, with listeners' choice of backchannels shaping narratives (e.g. 'mm' or 'crikey!' (1):5152) (Bavelas et al., 2000; Tolins and Fox Tree, 2014).

As seen in (1):5148, backchannels do not just occur at the ends of sentences or turns, but can occur subsententially. Grounding thus occurs incrementally, before a complete proposition has been produced or processed. Despite this, evidence suggests that there are places within and between turns where backchannels are salient. These *backchannel relevance spaces* (BRSs: Heldner et al., 2013), are analogous to but more common than transition relevance places (TRPs) – places where the turn may shift between speakers (Sacks et al., 1974). Feedback is optional at these points, and there are many reasons why any given BRS might not contain a backchannel or other feedback (e.g. individual variation). There may also be subtle nonverbal feedback which further complicates efforts to automatically predict where backchannels occur in dialogue. For practical dialogue systems, the positioning of backchannels is crucial. However although using low-level features (Cathcart et al., 2003; Gravano and Hirschberg, 2009) may allow a dialogue model to sound 'more human', it can't provide any insight into why feedback occurs where it does. Further, models in which feedback incorporates reasoning about the intentions or goals of one's interlocutor (Visser et al., 2014; Buschmeier and Kopp, 2013; Wang et al., 2011) presuppose a level of complexity that is unnecessary in natural conversation (Gregoromichelaki et al., 2011).[3]

[1] Examples are all taken from dialogue KB2 in the British National Corpus (BNC: Burnard, 2000).

[2] Although we believe that our analysis also applies to non-verbal feedback, in this paper we focus on verbal feedback.

[3] We are not claiming that people never use higher level reasoning – both in terms of general dialogue or in terms of appropriate backchannel placement – just that it is not necessary that they do so. This is especially clear from dialogues with young children who do not yet have higher-level mind reading skills, but still produce appropriate backchannel behaviour. We

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. Evidence 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.

(1)

A	5143	He did mashed potatoes
J	5144	Mm.
A	5145	cabbage, savoy cabbage, carrots ⟨pause⟩ and he'd cu- cut them like I always cut them cos they were only them little baby carrots so, what I do I slice them down
J	5146	Yeah.
A	5147	you know, down middle like
J	5148	Yeah.
A	5149	into quarters so I do them longer
J	5150	Yeah.
A	5151	and he'd done them like that in microwave for eight minutes ⟨pause⟩ and er, done sprouts ⟨pause⟩ then he'd put this meat pie in oven
J	5152	Crikey!
A	5153	and er, done onion gravy!
J	5154	Mm mm!
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.
A	5161	he says, I put a bit a ⟨pause⟩ Italian mixed herbs in middle of meat pie in my hand, put [them]
J	5162	[Mm.] [⟨laugh⟩]
A	5163	[and a bit] of Bovril.

2 Modelling feedback relevance spaces (FRSs)

In this section we briefly outline the formal tools used. Eshghi et al. (2015) provide a low-level semantic model of feedback integration in dialogue, and here we extend the model to explain why feedback tends to occur at certain points in an utterance. The model accounts for cases where feedback occurs at FRSs, and also provides an account of how feedback at inappropriate points can be integrated or interpreted.

2.1 Dynamic Syntax and Type Theory with Records

Dynamic Syntax (DS: Kempson et al., 2001; Cann et al., 2005) is an action-based grammar formalism, which models the word-by-word incremental processing of linguistic input. DS models the linear construction of *interpretations* without an independent level of syntactic representation, such that the output for any given string of words is a semantic tree representing predicate/argument structure. DS lends itself to the analysis of dialogue (Purver et al., 2006; Kempson et al., 2016, a.o.) as there is no stipulation for a separate parsing/production module, and speaker and hearer actions are the same; except that the current speaker has a more advanced goal-tree that subsumes the current tree. Recently, DS has been integrated

acknowledge that this means that models that include extra features such as acknowledger confidence (Visser et al., 2014) may capture an additional level of complexity over and above our notion of where backchannels are semantically licensed – indeed, the inclusion of such features may help explain when 'infelicitously' placed backchannels are interpretable (see Section 2.4, below) or what type of backchannel is more appropriate at a certain point.

with Type Theory with Records (TTR: Cooper, 2005) to provide the formalism in which semantic representations are couched (Eshghi et al., 2012; Eshghi, 2015; Purver et al., 2011) – see e.g. Fig. 1. TTR, with its rich notions of underspecification and sub-typing has proven crucial in (1) subsentential, incremental specifications of utterance content; (2) specifications of richer notions of dialogue context (Purver et al., 2010; Ginzburg, 2012); and (3) models of grammar learning and dialogue systems (Eshghi et al., 2013; Eshghi and Lemon, 2014; Kalatzis et al., 2016).[4]

Tree nodes in DS-TTR correspond to terms in the lambda calculus, decorated with labels expressing their semantic type and semantics; beta-reduction determines the type and formula at a mother node from those at its daughters (Figure 1). Trees can be *partial*, with unsatisfied *requirements* (e.g. $?Ty(e)$ is a requirement for development to $Ty(e)$) and contain a *pointer*, $\diamond$, labelling the node under development. Grammaticality is defined as processability in a context: the successful incremental construction of a tree with no outstanding requirements using all information given by the words in a string.

The parsing process 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 induced by parsing particular lexical items. 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 (e.g. see Fig. 3).

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 COMPLETION, which moves the pointer up and out of a sub-tree once all requirements therein are satisfied. Others reflect the fundamental predictivity and dynamics of DS. These apply optionally whenever their preconditions are met, but are not triggered by lexical input. The successful parse of a word w_1 amounts to finding a sequence of computational actions (possibly empty) that leads to a tree which satisfies the preconditions of the lexical actions for w_1. The parse search process/history can thus be represented as a Directed Acyclic Graph (DAG), with (partial) semantic trees as nodes, and actions as edges, i.e. transitions between trees.

Fig. 1 shows "John arrives", parsed incrementally, starting with the axiom tree, T_0, and ending with a complete tree. The intermediate step shows the effect of COMPLETION, which moves the pointer up and out of a complete node - this process is central in our explanation of FRSs.

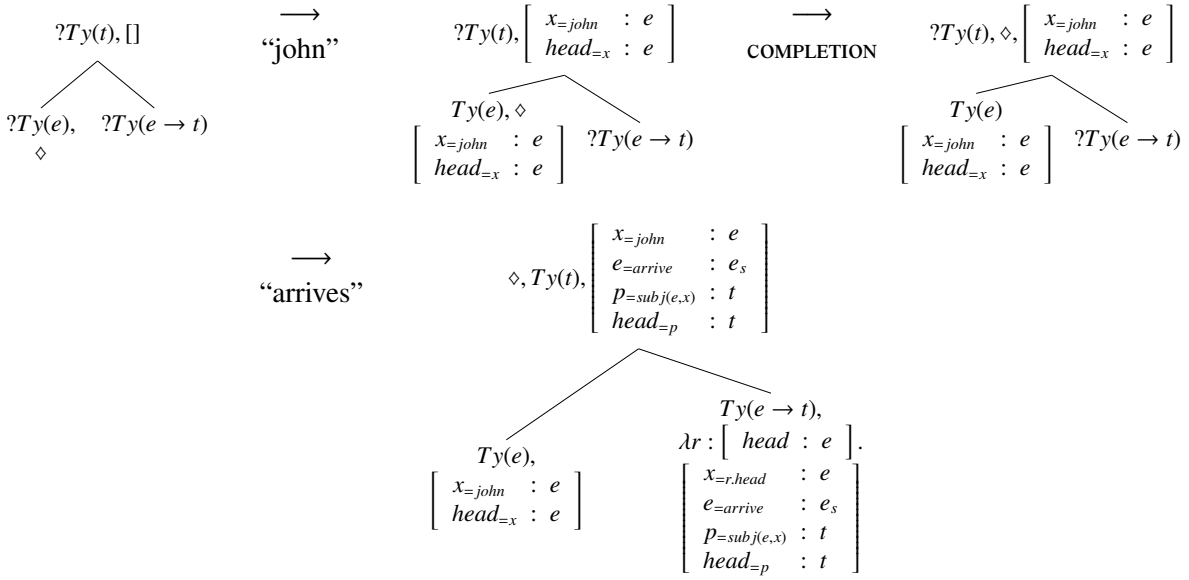

Figure 1: Incremental parsing in DS-TTR: *"John arrives"*

[4]Implementation of DS-TTR and the feedback model: `https://bitbucket.org/dylandialoguesystem/`

2.2 Context and the integration of feedback in DS

In DS, context, required for processing various forms of context-dependency – including pronouns, VP-ellipsis, self-repair and short answers – is the parse search DAG (Sato, 2011; Eshghi et al., 2012; Kempson et al., 2015). We take a coarse-grained view of the DAG with edges corresponding to words (sequences of computational action followed by a lexical action) rather than single actions, and dropping abandoned parse paths (see Hough, 2015, for details) - Fig. 2 shows an example.

As Eshghi et al. (2015) show, grounding (the integration into context of positive and negative feedback) can be captured using the context DAG, augmented with two *coordination pointers*: the *self-pointer*, ♦; and the *other-pointer*, ◇, marking where the speaker and hearer have each reached. Any utterance causes DAG pointer movement: the self-pointer tracks where the speaker has got to in production, and the other-pointer tracks where the listener has given feedback for reaching. This model accounts for negative and positive feedback. Negative feedback, e.g. clarification requests, causes branching in the DAG, where the current path is abandoned and another branch constructed – subsequent positive feedback realigns the two pointers. Contrarily, positive feedback e.g. backchannels and utterance continuations do not create new branches, but move the other-pointer forward on the current path.

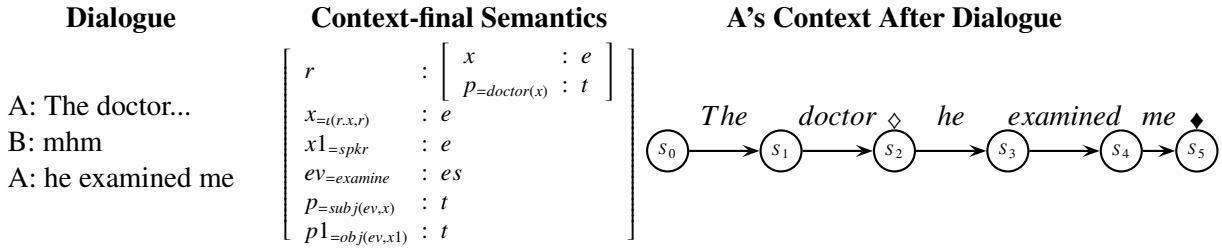

Figure 2: Backchannels as movement of context DAG coordination pointers. From A's perspective.

Fig. 2 shows a simple example of how a backchannel is integrated: this is A's context after processing the dialogue. After producing the first utterance, A's self-pointer, ♦, is on s_2, the right-most node of the DAG so far. B's backchannel provides positive feedback, thus moving A's other-pointer, ◇, to the same node, 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 feedback from B to be grounded: divergence of pointer positions thus represents 'forward momentum' in conversation. According to this model, the intersection of the path back to root from the self- and other-pointers is taken to be grounded.

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 no semantic content, to the strongest – utterance continuations, all serve to narrow down the otherwise mushrooming space of possible processing paths for interpretation. Their pervasiveness in dialogue is therefore not coincidental, but strategic, and serves to make interpretation in dialogue *locally* computationally tractable.

2.3 Feedback Relevance Spaces

Extending the model in Eshghi et al. (2015), we take backchannels to signal (when produced), or trigger (when parsed) COMPLETION. Feedback is therefore salient when a sub-tree has just been completed, i.e. when a semantic functor and argument have *both* been processed; for example, a backchannel is not normally licensed after a determiner (before the noun has been encountered) or after e.g. "John went to".

Fig. 3 shows the lexical entry for backchannels. This action ensures that after parsing a backchannel, we end up with a maximally complete tree. Specifically, for a backchannel to be parsable, we should not be on a node that is type-incomplete or on a node that is complete when its sister node is also complete - i.e. if more tree completion (BETA-REDUCTION, THINNING and COMPLETION) can still be done. This mechanism has been tested in the Dynamic Syntax implementation (Eshghi et al., 2011; Eshghi, 2015).

	IF	$?Ty(X)$	
	THEN	`abort`	
	ELSE	IF	$\langle\uparrow_0\downarrow_1\rangle\exists x.Tn(x)$
mhm			$\langle\uparrow_0\downarrow_1\rangle\neg\exists x.?x$
			$\neg\exists x.?x$
		THEN	`abort`
		ELSE	`do-nothing`

Figure 3: Lexical Entry for a backchannel

This allows us to explain feedback that comes after a semantic unit of information, thus grounding it. Further, it predicts that parse paths leading to further qualification become less likely when followed by a backchannel (e.g. in "A: Matt, B: mmm", A is less likely to further qualify/extend 'Matt'). A lack of feedback at such points is interactionally relevant. Elaboration should be more likely if no feedback is provided (e.g. A: Matt [no feedback], my brother ...). In dialogue, speakers often prompt or look for feedback using non-verbal behaviours such as gaze (Hjalmarsson and Oertel, 2012) and intonation (Gravano and Hirschberg, 2009). We hypothesise that this type of cue should occur at (or just before) points in the dialogue where Completion can occur.

2.4 Feedback at non FRSs

In our model, there are two possibilities for when feedback is produced at a point where it ought not be appropriate. 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) – i.e. moving ones other-pointer to the most recent position in the DAG at which Completion could have occurred.

The second, more interesting, possibility is where feedback is produced early. In this case, feedback seemingly precedes the completion of a semantic unit, which ought to be impossible. However, early feedback may be licensed where the completion is highly predictable. This is due to two key components of DS; i) predictability, which comes about from lexical and computational actions that induce more tree structure with requirements for fixed decorations as well as the reuse of actions and ii) the parity between parsing and production in DS. As shown in DS accounts of cross-person completions (Purver et al., 2010; Eshghi et al., 2012), 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. We propose that exactly the same mechanisms are exploited in cases of early feedback; in (2):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. Similarly, in (1):5158, J can produce a completion of A's prior turn that also functions as a clarification request. This analysis is supported by a text chat experiment in which producing candidate completions as clarifications turns out to be a fairly common strategy in response to artificially truncated turns – particularly when the part of speech of the upcoming material is predictable, and the context is sufficiently constrained (Howes et al., 2012).

(2)

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]

3 Conclusions

We have presented an analysis of feedback in dialogue using DS-TTR, which unifies the dialogue phenomena of backchannels, clarifications and completions in terms of their grounding actions. All these phenomena occur subsententially, and serve to signal how the dual processes of divergence and convergence that are crucial to successful interaction are managed locally, in a way that makes the search space tractable at a given point in an exchange.

We have hypothesised that FRSs are interactionally relevant parts of a utterance, analogous to TRPs, and that these are based on low-level semantic criteria. Further, we speculate that when feedback such as backchannels occurs at points other than FRSs it gets interpreted as if it had done so – either because the feedback is late and grounding the previous informational unit, or because it's early and the (rest of the) informational unit is predictable.

We have provided a precise, formal model of backchannels, their licensing, and effect. The evidence presented in this paper, while consistent with our model, is thus far circumstantial. We are therefore planning some corpus and experimental studies that directly bear on the clear empirically testable predictions provided by our model.

This work has implications for the production and interpretation of human-like feedback in dialogue systems; not just based on unanalysed features (which may result in accurate placement), but because they have successfully compiled a semantic unit at the point at which they produce or parse feedback.

Acknowledgements

Work on this paper was supported by two project grants: *Incremental Reasoning in Dialogue (IncReD)* VR (2016-01162); and *Babble: Domain-general methods for learning natural spoken dialogue systems* EPSRC (EP/M01553X/1).

References

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

Bavelas, J. B., P. De Jong, H. Korman, and S. S. Jordan (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*.

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.

Burnard, L. (2000). *Reference Guide for the British National Corpus (World Edition)*. Oxford University Computing Services.

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., 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 tenth EACL conference*, pp. 51–58. Association for Computational Linguistics.

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

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

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

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)*.

Eshghi, A., J. Hough, and M. Purver (2013). 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., J. Hough, M. Purver, R. Kempson, and E. Gregoromichelaki (2012). Conversational interactions: Capturing dialogue dynamics. In S. Larsson and L. Borin (Eds.), *From Quantification to Conversation: Festschrift for Robin Cooper on the occasion of his 65th birthday*, Volume 19 of *Tributes*, pp. 325–349. London: College Publications.

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

Eshghi, A. and O. Lemon (2014). How domain-general can we be? Learning incremental dialogue systems without dialogue acts. In *Proceedings of Semdial 2014 (DialWatt)*.

Eshghi, A., M. Purver, and J. Hough (2011). Dylan: Parser for dynamic syntax. Technical report, Queen Mary University of London.

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

Gravano, A. and J. Hirschberg (2009). Backchannel-inviting cues in task-oriented dialogue. In *INTERSPEECH*, pp. 1019–22.

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

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

Hjalmarsson, A. and C. Oertel (2012). Gaze direction as a back-channel inviting cue in dialogue. In *IVA 2012 workshop on Realtime Conversational Virtual Agents*, Volume 9.

Hough, J. (2015). *Modelling Incremental Self-Repair Processing in Dialogue*. Ph. D. thesis, Queen Mary University of London.

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

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

Kalatzis, D., A. Eshghi, and O. Lemon (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., T. Yamaguchi, K. Inoue, K. Takanashi, and N. Ward (2016). Prediction and generation of backchannel form for attentive listening systems. In *Proc. INTERSPEECH*, Volume 2016.

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

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

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

Poppe, R., K. P. Truong, and D. Heylen (2011). Backchannels: Quantity, type and timing matters. In *International Workshop on Intelligent Virtual Agents*, pp. 228–239. Springer.

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

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., E. Gregoromichelaki, W. Meyer-Viol, and R. Cann (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., C. Howes, E. Gregoromichelaki, and P. G. T. Healey (2009, September). Split utterances in dialogue: A corpus study. In *Proceedings of the 10th Annual SIGDIAL Meeting*, London, UK, pp. 262–271. Association for Computational Linguistics.

Sacks, H., E. Schegloff, and G. Jefferson (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, and C. Howes (Eds.), *The Dynamics of Lexical Interfaces*. CSLI Publications.

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

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.