How mechanistic can accounts of interaction be?

Ruth Kempson[†], Eleni Gregoromichelaki[†], Andrew Gargett[†],Chris Howes[‡], Greg Mills[‡], Matt Purver[‡]
[†]King's College London, Strand, London WC2R 2LS, UK
{ruth.kempson,eleni.gregor, andrew.gargett}@kcl.ac.uk
[‡]Queen Mary University of London, Mile End Road, London E1 4NS, UK
{chrizba,gj,mpurver}@dcs.qmul.ac.uk

Abstract

Ever since dialogue modelling first developed relative to broadly Gricean assumptions about utterance interpretation (Clark, 1996), it has been questioned whether the full complexity of higher-order intentions is made use of in everyday conversation. In this paper, building on the DS account of *split utterances*, we seek to further undermine the necessity of full-intention recognition/formation by exploring the extent to which the interactive coordination of dialogue exchange can be seen as emergent from mechanisms of language processing without needing representation by interlocutors of each other's mental states, or fully developed intentions as regards messages to be conveyed, even in e.g. clarifications and completions when the content of the utterance is in doubt.

1 Introduction

The pioneering work of H. Clark (Clark, 1996) initiated a broadly Gricean program for dialogue modelling with successful coordination in communication achieved through the presupposition of mutual beliefs constituting the participants' common ground and the establishment of joint intentions as a precondition for successful dialogue exchanges. However, computational models in this vein have very largely been developed without explicit high-order meta-representations of other parties' beliefs or intentions, except where dealing with highly complex dialogue domains (e.g. non-cooperative negotation (Traum et al., 2008)) or phenomena (e.g. collaborative completions (Poesio and Rieser, to appear)). With concepts such as *dialogue game board*, *QUD*, (Ginzburg, 1995; Larsson, 2002) *settledness* (Asher and Gillies, 2004) largely replacing intention recognition, it is arguable that Gricean assumptions underpinning communication should be re-considered. A parallel weakening has been taking place within another major pragmatic paradigm, that of (Sperber and Wilson, 1986). The relevance-theoretic view is that the content of an utterance is established by a hearer relative to what the speaker could have intended (relative also to a concept of mutual manifestness of background assumptions). However, (Breheny, 2006) argued that children in the initial stages of language acquisition communicate relative to a weaker 'naive-optimism' view in which some context-established interpretation is simply presumed to match the speaker's intention, only coming to communicate in the full sense substantially later (see (Tomasello, 2008) for a Gricean variant of this view).

With this weakening across all pragmatic models of the status of recognition of other interlocutor's intentions, for at least some cases of communication, in this paper we set out the groundwork for an interactive model of communication using *Dynamic Syntax* (*DS*: (Cann et al., 2005)), and examine its application to the tightly interactive dialogue phenomena that arise in cases of continuation/clarification/reformulation splits between speakers. In this model, each party to the dialogue interprets the signals they receive, or plans the signals they send, relative to their own context, without explicit (meta-)representation of the other party's knowledge/beliefs/intentions. Nevertheless the effect of coordinated communication is achieved by relying on ongoing feedback between parties and the goal-directed action-based architecture of the grammar.

Our claim is that in all cases where any single agent's system fails to fully determine choices to be made (either in parsing or production), the choice made may happen to be right, and might or might not get acknowledgement; it may be wrong and get corrected, thereafter establishing explicit coordination with respect to some subpart of the communication; or, in recognition of the non-determinism, the agent may set out a sub-routine of clarification thereby delegating the choice of construal to the interlocutor before proceeding. Success in communication thus involve clarifi-

cation/correction/extension/reformulation ("repair strategies") as essential subparts of the exchange. Pursued non-incrementally, such strategies may induce a sense of non-monotonic repair and the need to revise established context. But pursued incrementally within a goal-directed architecture, these do not constitute communication breakdown and repair, but the normal mechanism of hypothesised update, context selection, and confirmation. By building on the assumption that successful communication may involve subtasks of repair,[1] the mechanisms for informational update that effect such interaction can be defined without any reliance on (meta-)representing contents of the interlocutors' mental states.

2 Split Utterances

Switching of roles between speaking and hearing, across and within sentential structures, is characteristic of dialogue. Participants are able to abandon one role and take up the other without any indication of this being problematic, even mid-utterance:

(1) Daughter: Oh here dad, a good way to get those corners out
Dad: is to stick yer finger inside.
Daughter: well, that's one way. [from Lerner (1991)]

(2) A: They X-rayed me, and took a urine sample, took a blood sample. Er, the doctor
B: Chorlton?
A: Chorlton, mhm, he examined me, erm, he, he said now they were on about a slight [shadow] on my heart. [BNC: KPY 1005-1008]

(3) A: Are you left or
B: Right-handed.

The challenge of modelling such phenomena has recently been taken up by (Poesio and Rieser, to appear) for German, defining a fine-grained neo-Gricean model of dialogue interactivity that builds on an LTAG grammar base. Their primary aim is to model *completions*, as in (1) and (3), with take-over by the hearer because the REMAINDER of the utterance is taken to be understood or inferrable from mutual knowledge. Their account hinges on two main areas: the assumption of recognition of interlocutors' intentions according to shared joint plans, and the use of incremental grammatical processing based on LTAG. However, the extent to which their account successfully captures such incremental switch-over turns on the assumption of a STRING-BASED level of syntactic analysis, for it is this which provides the top-down, predictive element allowing the incremental integration of such continuations. The question we address here is whether the sparer DS grammar, dispensing with an autonomous syntax, can provide the required *predictivity* (for this psycholinguistic notion, see (Sturt and Crocker, 1996)); and indeed, besides its greater economy in representational levels, such a model seems better suited to capturing such phenomena since there are cases which do NOT involve interlocutors intending to say the same string of words/sentence:

(4) with smoke coming from the kitchen:
A: Have you burnt the
B buns. Very thoroughly.
A: But did you
B: burn myself? No. Luckily.

The explanation for B's continuation in the fourth turn of (4) cannot be string-based as then *myself* would not be locally bound (its antecedent is *you*). Moreover, in LTAG words are defined in terms of syntactic/semantic pairings, relative to a given head, with adjuncts as a means of splitting these. However, as all of (1)-(4) show, utterance take-over can take place at ANY point in a sequence of words, with or without a head having occurred prior to the split. And many split utterances are not joint SENTENTIAL constructions; as (2),(3) show, even the function of the utterance (e.g. statement/clarification, question/answer) can alter in the switch of roles, with fragments playing multiple roles (e.g. question/completion/acknowledgment/answer) at the same time. If the grammar necessarily induces speech act representations these cannot be accounted for but as cases of ambiguity or requiring hidden constituent reconstruction.

The setting for the Poesio and Rieser analysis is one in which participants are assigned a collaborative task with a specific joint goal, so joint intentionality is fixed in advance and hence anticipatory computation of interlocutor's intentions can be defined. However, (Mills and Gregoromichelaki, 2008in prep) argue that even in such task-specific situations joint intentionality is not guaranteed but

[1] This view is notably not dis-similar to the view of (Ginzburg, forthcmg)

rather has to evolve as a result of routinisation. In accordance with this, as (1) shows, in ordinary conversation, there is no guarantee that there is a plan genuinely shared, or that the way the shared utterance evolves is what either party had in mind to say at the outset, indeed obviously not, as otherwise such exchanges would appear otiose. Instead utterances are shaped incrementally according to feedback by the interlocutor (CA ref here). And, as in (2), clarification can occur well before the completion of the utterance, which then absorbs both contributions. Grammatical integration of such joint contributions must therefore be flexible enough to allow such switches, with fragment resolutions occurring incrementally before computation of intentions at the pragmatic level is even possible.

The Poesio and Rieser account marks a significant advance in the analysis of such phenomena as it employs an incremental view of the grammar in their analysis. But, as we saw above, the phenomenon is more general than just *completions*, the primary target of the Poesio and Rieser account. Moreover, given the observations above, dialogue exchanges involving incremental split utterances of any type are hard to model adopting any other static grammatical framework. First of all, in such frameworks it is usually the sentence/proposition that is the unit of syntactic/semantic analysis, and, in the absence of an incremental/parsing perspective, elliptical phenomena/fragments are defined (following (Dalrymple et al., 1991)) as associated with an abstraction operation over contextually provided propositional content to yield appropriate functors to apply to the fragment. But this problematically increases parsing uncertainty, since multiple options of appropriate "antecedents" for elliptical fragments become available (one for each available abstract). In consequence, to resolve such exploding ambiguities, the parsing mechanism has to appeal to general pragmatic mechanisms having to do with recognising the speaker's intention in order to select a single appropriate interpretation. The conundrum that opens up is that intention recognition on which all such successful contextual resolution will have to be based is inapplicable in such sub-sentential split utterances in all but the most task-specific domains, for, in principle, attribution to any party of specific recognition of the speaker's intention to convey some specific propositional content is unavailable until some propositional formula is established. Recognition of such fully propositional intentions thus cannot be the basis on which incrementally established joint utterances are based. Moreover, from a generation point of view relative to orthodox grammar-producer assumptions, the fact that speakers are interrupted, with (possibly unintended) continuations of their utterances being provided instead, means that the original speaker's plan to convey some full proposition will have to be abandoned mid-production, with some form of radical revision initiated in adopting the role of the parser. However, the seamlessness of such switches indicates no radical revision, as is confirmed by psycholinguistic evidence that speakers do not start articulating with fully formed propositional contents to convey already in mind ((Fernanda and Swets, 2002)).

Below we set out a model of parsing and production mechanisms that make it possible to show how, with speaker and hearer in principle using the same mechanisms for construal, equally incrementally applied, issues about interpretation choice and production decision may be resolvable without reflections on the other party's mental state but solely on the basis of feedback. As we shall see, what connects our diverse examples, and indeed underpins the smooth shift in the joint endeavour of conversation, lies in *incremental*, context-dependent processing and tight coordination between parsing and generation, essential ingredients of the DS dialogue model ((Cann et al., 2005)). Instead of data such as (1)-(4) being problematic for such an account, in fact, their extensive use illustrates the advantages of a DS account in its provision of restricted contextually salient structural frames within which fragment construal/generation take place, narrowing down the threatening multiplicity of interpretations by incrementally weeding out possibilities en route to some commonly shared understanding. Features like incrementality, *predictivity* and context-dependent processing are, that is, built into the grammar architecture itself: EACH successive processing step relies on a grammatical apparatus which integrates essential reference to the context in order to proceed. Such a view notably does not invoke high-level decisions about speaker/hearer intentions as part of the mechanism itself. That this is the right view to take is enhanced by the fact that, as all of (1)-(4) show, neither party in such role-exchanges can definitively

know in advance what will emerge as the eventual joint proposition. An additional puzzle for any common-ground/intention-based views is that both speakers and hearers may elect not to make use of what is well established shared knowledge. On the one hand, in selecting an interpretation, a hearer may fail to check a putative interpretation against consistency with what they believe the speaker could have intended (as in (5) where B construes the fragment in flagrant contradiction to what she knows B knows):

(5) A: Why don't you have cheese and noodles?
B: Beef? You KNOW I'm a vegetarian

On the other hand, speaker's choice of anaphoric expression, supposedly restricted to well-established shared knowledge, is commonly made in apparent neglect of their hearer (CAN WE FIND A CORPUS EG):

(6) having read out newspaper headline about Brown and Obama, speaker upon reading next headline provides as follow-on:
A: They've received 10,000 emails.
B: Brown?
A: The Camerons.

Given this type of example, checking in parsing or producing utterances whether information is jointly held by the dialogue participants, the (perceived) *common ground*, can't be a necessary condition on such activities. Hence it is not intrinsic to utterance interpretation in virtue of which conversational dialogue takes place. So we turn to Dynamic Syntax (DS) to explore possible forms of correlation between parsing and generation as they take place in dialogue without reliance on any such construct.

3 Incrementality in Dynamic Syntax

DS is a procedure-oriented framework, involving incremental processing, i.e. strictly sequential, word-by-word interpretation of linguistic strings. The notion of incrementality in DS is closely related to another of its features, the *goal-directedness* of BOTH parsing and generation. At each stage of processing, *structural predictions* are triggered that could fulfill the goals compatible with the input, in an underspecified manner. For example, when a proper name like *Bob* is encountered sentence-initially in English, a semantic predicate node is predicted to follow ($?Ty(e \rightarrow t)$), amongst other possibilities.

By way of rehearsing DS devices, let us look at some formal details with an example, *Bob saw Mary*. The 'complete' semantic representation tree resulting after the complete processing of this sentence is shown in Figure 1 below. A DS tree is formally encoded with the tree logic *LOFT* (Blackburn and Meyer-Viol, 1994), is generally binary configurational, with annotations at every node which represent semantic formulae with their type information (e.g. '$Ty(x)$') based on a combination of the epsilon and lambda calculi.

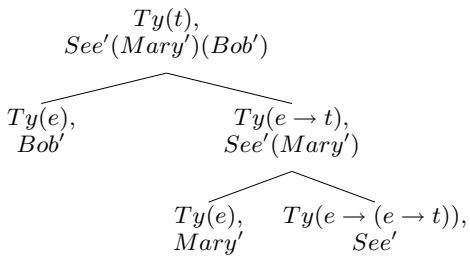

Figure 1: A DS complete tree

Such complete trees are constructed, starting from a radically underspecified goal, the *axiom*, the leftmost minimal tree in the illustration provided by Figure 2. Going through *monotonic updates* of partial or *structurally underspecified* trees, complete trees are eventually achieved. Crucial for expressing the goal-directedness are *requirements*, i.e. unrealised but expected node/tree specifications, indicated by '?' in front of annotations. The axiom says that a proposition (of type *t*, $Ty(t)$) is expected to be constructed. Furthermore, the *pointer* notated with '$\Diamond$' indicates the 'current' node in processing, namely the one to be processed next, and governs word order.

Updates are carried out by means of applying *actions* of two types. *Computational actions* govern general tree-constructional processes, such as moving the pointer, introducing and updating nodes, compiling interpretation for all non-terminal nodes. In Figure 2, the update of (1) to (2) is executed via computational actions specific to English, expanding the axiom to the subject and predicate nodes, requiring the former to be processed next. Construction of only weakly specified tree relations (*unfixed nodes*) can also be induced, characterised only as dominance by some current node, with subsequent update required. Individual lexical items also provide procedures for building structure in the form of *lexical actions*, inducing both nodes and annotations. In the update from (2) to (3), the set of lexical

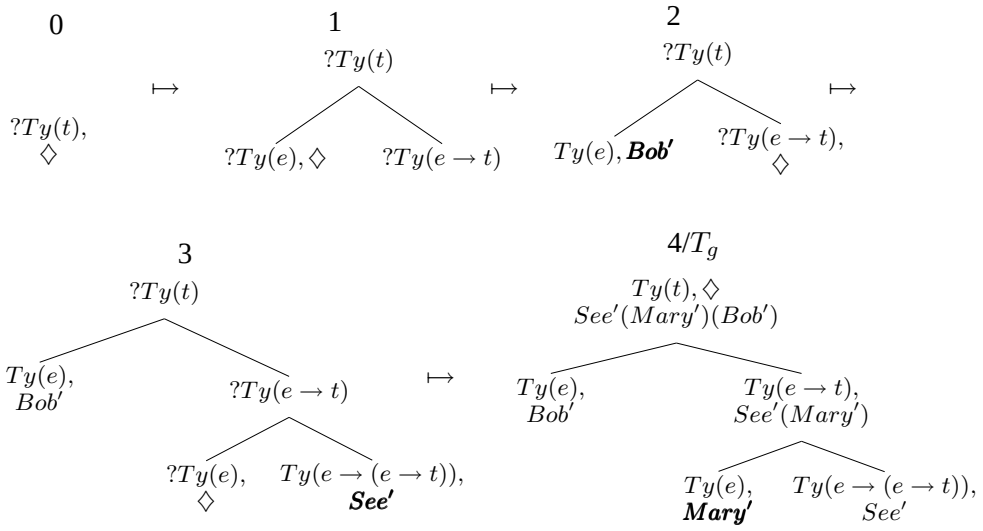

Figure 2: Monotonic tree growth in DS

actions for the word *see* is applied, yielding the predicate subtree and its annotations. Unlike conventional bottom-up parsing, the DS model takes the parser/generator to entertain some predicted goal(s) (*requirements*) to be reached eventually at any stage of processing. Thus *partial trees* grow incrementally, driven by procedures associated with particular words as they are encountered.

Individual DS trees consist of predicates and their arguments. Complex structures are obtained via a general tree-adjunction operation licensing the construction of so-called *linked* trees, pairs of trees where sharing of information occurs. The assumption in the construction of such linked structures is that at any arbitrary stage of development, some type-complete subtree may constitute the context for the subsequent parsing of the following string as an adjunct structure candidate for incorporation into the primary tree, hence the obligatory sharing of information in the resulting semantic representation. Appositional structure can then be established by defining a link transition as in Figure 3 from a node of type e in which a skeletal epsilon term[2] has been constructed (with all terminal nodes decorated but nonterminals not fully compiled) onto a linked tree introduced with a requirement to develop a term using that very same variable.

A twinned evaluation rule then combines the restrictors of two such paired terms to yield a composite term (unlike the Poesio and Rieser account, this does not involve ambiguity of the head NP according to whether a second or subsequent NP follows). The fact that the first term has not been completed is no more than the term-analogue of the delaying tactic made available by expletive pronouns, allowing term modification when the pointer returns from its sister node immediately prior to compiling the decorations of its mother (as in *A man has won the lottery, someone you know*), and providing motivation for asking for clarification if completion without such clarification isn't possible.

Such linked trees and their development set the scene for a general characterisation of context. *Context* in DS is defined as the storage of *parse states*, i.e., the storing of partial tree, word sequence parsed to date, plus the actions used in building up the partial tree. All fragments illustrated above are processed by means of either extending the current tree, or by constructing linked structures with transfer of information among them so that one tree provides the context for another. Such fragments are licensed as well-formed by the grammar only relative to such contexts (Gargett et al., 2008; Kempson et al., 2009).

4 Parsing/Generation Coordination

This architecture allows a dialogue model in which generation and parsing function in parallel, following exactly the same procedure in the

[2] *Epsilon terms*, like $\epsilon, x, Consultant'(x)$, stand for witnesses of existentially quantified formulae in the epsilon calculus and represent the semantic content of indefinites. Defined relative to the equivalence $\psi(\epsilon, x, \psi(x)) = \exists x \psi(x)$, their defining property is their reflection of their containing environment, and accordingly they are particularly well-suited to expressing the growth of terms secured by such appositional devices.

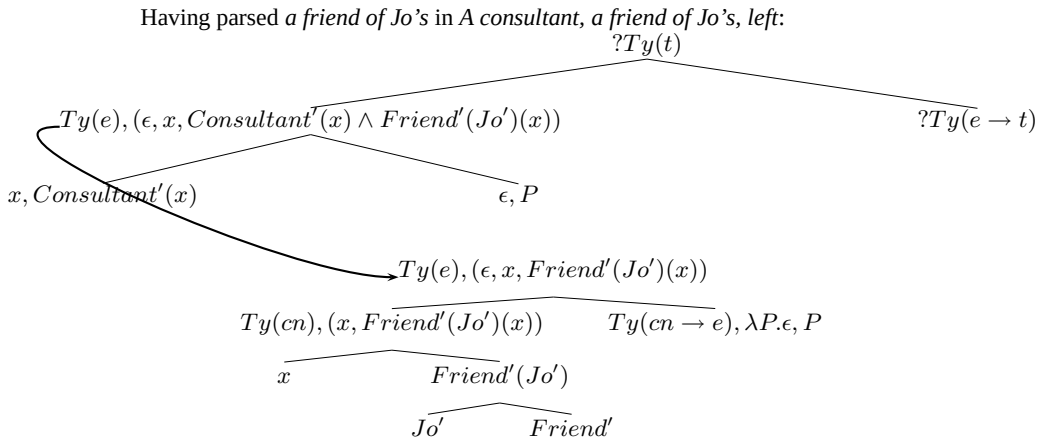

Figure 3: Apposition in DS

same order. Figure 2 also displays the generation steps 0 to 4 of *Bob saw Mary*, for generation of this utterance follows precisely the same actions and trees from left to right as in parsing, although the complete tree is available as a *goal tree* from the start (hence the labelling of the complete tree as T_g): in this case the eventual message is known by the speaker, though of course not by the hearer. What generation involves in addition to parse steps is reference to this goal tree to check whether each intended generation step (1, 2, 3, 4) is consistent with it. That is, a *subsumption* check is carried out as to whether the current parse tree is monotonically extendible to the goal tree. The trees (1-3) are licensed because each of these subsumes T_g in this sense. Each time then the generator applies a lexical action, it is licensed to produce the word that carries that action only under successful subsumption check: at Step (3), for example, the generator processes the lexical action which results in the annotation 'See'', and upon success and subsumption of T_g license to generate the word *see* ensues.

For processing split utterances, two more consequences are pertinent. First, there is nothing to prevent speakers initially having only a partial structure to convey, i.e. T_g may be a *partial* tree: this is unproblematic, as all that is required by the formalism is monotonicity of tree growth, and the subsumption check is equally well defined over partial trees. Second, the goal tree T_g may change during generation of an utterance, as long as this change involves monotonic extension; and extensions/clarifications/repairs within and across speakers in DS can be straightforwardly modelled by appending a linked structure projecting added material to be conveyed (preserving monotonicity):

(7) A friend is arriving, with my brother, maybe with a new partner.

Such a model under which the speaker and hearer essentially follow the same sets of actions, each incrementally updating their semantic representations, allows the hearer to 'mirror' the same series of partial trees as the producer, albeit not knowing in advance the content of the unspecified nodes. Furthermore, not only can the same sets of actions be used for both processes, but also a large part of the parsing and generation algorithms is shared. In principle both parties may engage with partial tree representations. Even the concept of *goal tree* may be shared between speaker and hearer, in so far as the hearer may have richer expectations relative to which the speaker's input is processed, as in the processing of a clarification question. Conversely, the speaker may have only a partial tree as a goal tree, relative to which they are seeking clarification. As no intervening level of syntactic structure over the string is computed, the parsing/generation tasks are more efficiently incremental in that semantic interpretation is imposed at each stage of lexical integration, irrespective of whether a given partially developed constituent is complete.

4.1 Split utterances in Dynamic Syntax

Split utterances follow as an immediate consequence of these assumptions. For the dialogues in (1)-(4), while A reaches a partial tree of what she has uttered through successive updates as described above, B as the hearer, will follow the

same updates to reach the same representation of what he has heard as they both apply the same parsing mechanism which is none other than their shared grammar.[3] This provides B with the ability at any stage to become the speaker, interrupting to continue A's utterance, repair, ask for clarification, reformulate, or provide a correction, as and when necessary. According to our model of dialogue repeating or extending a constituent of A's utterance by B is licensed only if B, the hearer now turned speaker, entertains a message to be conveyed (a new *goal tree*) that matches or extends the parse tree of what he has heard in a monotonic fashion. This message (tree) may of course be partial, as in (2) where B is adding a clarificational linked structure to a still-partially parsed antecedent. Importantly in DS, both A and B can now reuse the already constructed (partial) parse tree in their immediate context as a point from which to begin parsing and generation, rather than having to rebuild an entirely novel tree or subtree. In this sense, the most recent parse tree constitutes the most immediately available local "antecedent" for fragment resolution, for both speaker and hearer, hence no separate computation or definition of *salience* or speaker intention by the hearer is necessary.

As we saw, the hearer B may respond to what he has built up in interpretation, anticipating the verbal completion as in (1)-(3). This is facilitated by the general predictivity of the DS architecture since the parser is always predicting top-down goals (*requirements*) to be achieved in the next step. Such goals are what drives the search of the lexicon in generation so a parser who shifts to a successful lexicon search before processing the anticipated lexical input provided by the speaker can become the generator and take over. In (3), B is doing just that, using this anticipation as simultaneously a completion of A's utterance, an acknowledgment of his understanding of the question and of his taking it up, and as a direct form of reply. Any framework which relies on complete disambiguation of speaker's intention in order to resolve such fragments does not allow for such multiple functionality. Such fragments would have to be characterised as multiply ambiguous instead. But the ambiguity stance once more increases parsing uncertainty so that inferential pragmatic mechanisms (appealing to deciphering speakers' intentions) have to be invoked to select the appropriate update rules that should or should not apply at this juncture.

[3] A completely identical grammar is, of course, an idealisation but one which is harmless for current purposes

5 Summary Evaluation

What the availability of these DS derivations shows is how, with grammar mechanisms defined to induce tree growth used incrementally in both parsing and generation, the core activities in dialogue can all take place without any other-party representation at all. This then results in a view of communication that is not grounded in recognising speaker's intentions, hence can be displayed by both young children and adults equally. The two crucial properties are the intrinsic predictivity/goal-directedness in the formulation of the DS action-updates, and the fact that both parsing and production can have arbitrary partial goals, so that in effect both interlocutors are building structures in tandem. Because of the assumed partiality of goal trees, speakers do not have to be taken to have fully formed messages to convey at the beginning of the generation task but can instead rely on feedback to shape their utterance. As goal trees are expanded incrementally, completions by the other party can be monotonically accommodated even though they might not represent what the speaker would have uttered if not interrupted: as long as what emerges as the eventual joint proposition is some compatible extension of the original speaker's goal tree, it will be accepted as sufficient for the purposes to hand. Hence "repair" phenomena naturally emerge as "coordination devices" (Clark, 1996), devices exploiting mutually salient contexts for achieving coordination enhancement. And such jointly constructed content through cycles of "miscommunication" and "repair" is more securely coordinated (see e.g. (Healey, 2008)) and thus can form the basis of what each party considers shared cognitive context.

It might appear that the analysis faces the familiar exponential explosion of interpretations requiring the computation of speaker intentions on the basis of common ground, albeit at a sub-propositional level. However, on the incremental processing view developed here, on the one hand, such speaker intentions are not available at the relevant juncture and, on the other hand, speaker intentions might not have even be formed given the partiality of the goal trees. But with feedback able to be provided and accommodated at any

(sub-propositional) stage, the potential exponential explosion of interpretations can be kept firmly in check: structurally, such fragments can be integrated in the current partial tree representation directly (given the position of the pointer) so there is no structural ambiguity multiplication. What is notable though is that for any one such intermediate check point, equal use of tree construction processes by the generator means that consistency checking can remain internal to each interlocutor's system, the fact of their mirroring each other resulting in their being at the same point of tree growth. Even in the case of repairs, these are processed relative to their own set of trees (background knowledge) and not with respect to what they think the other might have in mind. This is compatible with a mechanistic view of dialogue processing as in (Pickering and Garrod, 2004) (although no invocation of "priming" mechanisms is involved).

Of course, DS being a grammar formalism, a full account of all facets of dialogue incorporating its non-monotonic aspects is not within its remit. Additionally, in principle, the account provided does not preclude the representation of intentions as explicitly expressed and manipulated (in the form of adjoined linked structures) derived through the mechanisms mentioned in (Poesio and Rieser, to appear) or an alternative routinisation account as preliminary indicated in (Mills and Gregoromichelaki, 2008in prep). Nonetheless, the dual applicability of the mechanisms defined identically for both parsing and (tactical) generation provides a basis for explaining how apparently shared contents can be incrementally established without either participant constructing hypotheses of what is entertained by the other, all context-based selections being based on the individual's own context as far as fragment resolution is concerned. Where uncertainty arises the context-dependent repair mechanisms can take over. This, in its turn, makes possible an account of how hearers may construct interpretations that are transparently inconsistent with what both interlocutors know ((5)-(6)). Hence we suggest, contra (Tomasello, 2008), we need to be exploring accounts of human communication as an activity involving emergent agent coordination without any necessary grounding in high-level mind-reading.

References

Nicholas Asher and Anthony Gillies. 2004. Common Ground, Corrections and Coordination. *Argumentation*, 17:481–512.

Patrick Blackburn and Wilfried Meyer-Viol. 1994. Linguistics, logic and finite trees. *Bulletin of the IGPL*, 2:3–31.

Richard Breheny. 2006. Communication and folk psychology. *Mind & Language*, 21(1):74–107.

Ronnie Cann, Ruth Kempson, and Lutz Marten. 2005. *The Dynamics of Language*. Elsevier, Oxford.

Herbert H. Clark. 1996. *Using Language*. Cambridge University Press.

Mary Dalrymple, Stuart M. Shieber, and Fernando C. N. Pereira. 1991. Ellipsis and higher-order unification. *Linguistics and Philosophy*, 14(4):399–452.

Ferreira Fernanda and Benjamin Swets. 2002. How incremental is language production? evidence from the production of utterances requiring the computation of arithmetic sums. *Journal of Memory and Language*, 46:5784.

Andrew Gargett, Eleni Gregoromichelaki, Chris Howes, and Yo Sato. 2008. Dialogue-grammar correspondence in dynamic syntax. In *Proceedings of the 12th SEMDIAL (LONDIAL)*.

Jonathan Ginzburg. 1995. Resolving questions, I. *Language and Philosophy*, 18(5):459–527.

Jonathan Ginzburg. forthcmg. *The Interactive Stance: Meaning for Conversation*. CSLI.

Patrick Healey. 2008. Interactive misalignment: The role of repair in the development of group sub-languages. In R. Cooper and R. Kempson, editors, *Language in Flux*. College Publications.

Ruth Kempson, Eleni Gregoromichelaki, and Yo Sato. 2009. Incrementality, speaker/hearer switching and the disambiguation challenge. In *Proceedings of European Association of Computational Linguistics proceedings.*

Staffan Larsson. 2002. *Issue-based Dialogue Management*. Ph.D. thesis, Göteborg University. Also published as Gothenburg Monographs in Linguistics 21.

Greg Mills and Eleni Gregoromichelaki. 2008/in prep. Coordinating on joint projects. Talk given at the Coordination of Agents Workshop, Nov 2008, KCL.

Martin Pickering and Simon Garrod. 2004. Toward a mechanistic psychology of dialogue. *Behavioral and Brain Sciences.*

Massimo Poesio and Hannes Rieser. to appear. Completions, coordination, and alignment in dialogue. Ms.

Dan Sperber and Deirdre Wilson. 1986. *Relevance: Communication and Cognition*. Blackwell.

Patrick Sturt and Matthew Crocker. 1996. Monotonic syntactic processing: a cross-linguistic study of attachment and reanalysis. *Language and Cognitive Processes*, 11:448–494.

Michael Tomasello. 2008. *Origins of Human Communication.* MIT.

David Traum, Stacy Marsella, Jonathan Gratch, Jina Lee, , and Arno Hartholt. 2008. Multi-party, multi-issue, multi-strategy negotiation for multi-modal virtual agents. In *8th International Conference on Intelligent Virtual Agents*, September.