Dialogue-Grammar Correspondence in Dynamic Syntax

Andrew Gargett[†], **Eleni Gregoromichelaki**[†], **Chris Howes**[‡], **Yo Sato**[‡]
[†]King's College London, Strand, London WC2R 2LS, UK
{andrew.gargett,eleni.gregor}@kcl.ac.uk
[‡]Queen Mary University of London, Mile End Road, London E1 4NS, UK
{chrizba,yosato}@dcs.qmul.ac.uk

Abstract

In this paper, we argue, contra a prevailing trend to classify elliptical structures in terms of sub-types specific to conversational dialogue, that despite their diversity of usage in conversational dialogue, such fragments are analysable in terms of structure-building mechanisms that have motivation elsewhere in the grammar (the framework adopted is *Dynamic Syntax*, Kempson et al. (2001); Cann et al. (2005)). The fragment types modelled include *reformulations*, *clarification requests*, *extensions*, *corrections* and *acknowledgements*. We go on to argue that the incremental use of such ellipses serves a specific role in dialogue, namely a means of incrementally narrowing down the range of otherwise mushrooming alternative structural and interpretative analyses, a problem known to constitute a major challenge to any parsing system. We conclude that with grammar seen as a set of parse procedures, we have a basis for an integrated characterisation of dialogue phenomena while nonetheless not defining a grammar of conversational dialogue.

1 Introduction

In confronting the challenge of providing formal models of dialogue, with its plethora of fragments and rich variation in modes of context-dependent construal, it might seem that linguists face two types of methodological choice. Either (a) conversational dialogue demonstrates dialogue-specific mechanisms, for which a grammar specific to such activity can be constructed; or (b) variation arises due to the employment of independent parsing/production systems which are nevertheless based on some mode-neutral grammar formalism. However, as dialogue research continues to develop, there are intermediate possibilities, and in this paper we discuss the approach developed within *Dynamic Syntax* (*DS*, Kempson et al. 2001, Cann et al. 2005), a grammar framework within which, not only the parser, but indeed "syntax" itself is seen as the progressive construction of semantic representations set in context. Here we extend the analyses presented in Kempson et al. (2007) to a range of further fragment types, in particular *reformulations*, *fragment requests* and *corrections* accompanied by *extensions*. From a DS perspective, such apparently dialogue-specific constructions can be seen to result from perfectly general structural processes, despite being characteristic of cross-party conversational data.

Further, we claim that the grammar itself constitutes the basis for parsing strategies that facilitate an efficient online processing, structural and semantic. In this respect, the DS dialogue model provides the means of achieving this *during* the course of the sub-sentential construction process, demonstrating that timely application of such generally available "syntactic" mechanisms directly contributes to the human processor's high degree of success in linguistic interaction. We conclude that, contrary to conventional assumptions of the grammar-parser feeding relation whereby the parser exclusively handles disambiguation, grammars, as employed in dialogue, can also be seen as contributing the mechanisms for restricting available interpretations provided their formal specification can be made to reflect this incremental facilitating function.

2 Background

The data we focus on are non-repetitive fragment forms of acknowledgements, clarifications and corrections[1]:

(1) A: Bob left.
B: (Yeah,) the accounts guy.

(2)
A. They X-rayed me, and took a urine sample, took a blood sample.
A: 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.

(4) A: Bob left.
B: Rob?
A: (No,) (Bob,) the accounts guy.

Even though in the literature the NP fragments in (2) - (4) might be characterised as distinct constructions, they all illustrate how speakers and hearers may contribute, in some sense to be made precise, to the joint enterprise of establishing some shared communicative content, in what might be loosely called *split utterances*. And even (1), an *acknowledgement*, can be seen this way upon analysis: B's addition is similar to an afterthought *extension* added to A's fully sentential utterance. It can be seen in (2) that such joint construction of content can proceed incrementally: the *clarification request* in the form of a *reformulation* is provided by B and resolved by A within the construction of a single proposition. The attested example in (3) represents an intermediate case, in which the respondent realising what the question is provides the answer AS the *completion* of the initiator's question, so that the fragment serves simultaneously as question and answer. In (4) the fragment reply involves *correction*, with parties to the conversation confronting the need for negotiation as to whose information is more reliable before any coordination can be said to be achieved. Nevertheless such corrections can be also *extensions* in the above sense, taken as providing a single conjoined propositional content before the requisite coordination can be said to have been achieved.

It might seem that such illustration of diversity of fragment uses is ample evidence of the need for conversation-specific rules to be articulated as part of a grammar. Indeed, Fernández (2006) presents a thorough taxonomy, as well as detailed formal and computational modelling of *Non-sentential Utterances* (NSUs), referring to contributions such as (1) as *repeated acknowledgements* involving *reformulation*. Since such fragments require contextual information singling out a particular constituent of the previous utterance, Fernández models such constructions via type-specific "accommodation rules" which make a constituent of the antecedent utterance "topical". The semantic effect of the acknowledgement is then derived by applying an appropriately defined utterance type for such fragments to the newly constructed context. A distinct form of contextual accommodation is employed to model so-called *helpful rejection* fragments, as in (4) (without the reformulation), whereby a *wh*-question is accommodated in the context by abstracting over the content of one of the sub-constituents of the previous utterance. The content of the rejection is derived by applying this *wh*-question in the context to the content of the fragment.

The alternative explored here is whether phenomena such as (1)-(2), both of which are non-repetitive appositional next-speaker contributions, can both be handled using the same mechanisms for structure-building made available in the core grammar, without recourse to conversation-specific extensions of that grammar and contextual accommodation rules. Given that the range of interpretations these fragments receive in actual dialogue seem to form continua with no well-defined boundaries and mixing of functions (see (3)-(4)) we propose that the grammar itself simply provides the mechanisms for processing and integrating such fragments in the current structure whereas the precise contribution of such fragments to the communicative interaction can be calculated by pragmatic inferencing or, as seems most often to be the case, be left underspecified. The framework within which the explanation will be provided is Dynamic Syntax, in which the dynamics of how information accrues in language processing is the core of the syntactic explanation.

[1] henceforth, A female, B male

One bonus of the stance taken here is the promise it offers for elucidating the grammar-parser contribution to the disambiguation task. Part of the challenge of modelling dialogue is the apparent multiplicity of interpretive and structural options opened up during processing by the recurrent, often overlapping fragments as seen in (2) above. Thus, it might seem that the rich array of elliptical fragments available in dialogue adds to the complexity of the interpretive task, owing to their high degree of context-dependence (hence the need for accommodation and construction-specific interpretation rules). However, an alternative point of view is to see such phenomena as providing a window on how interlocutors exploit the incrementality of linguistic processing to manage the explosion of interpretative/structural options multiplying at each step. The context-dependent interpretation of fragments, when employed incrementally, enables the hearer to immediately respond to a previous utterance at any relevant point in the construction process, thereby enabling interlocutors to (incrementally) constrain interpretation during the very process in which it is developed.

Modelling this kind of flexibility in processing requires fine-grained control of how the current utterance can be combined with previous contextual information. Grammatical frameworks which take the radical context dependency of linguistic processing as being outside the remit of the grammar might make it seem that these phenomena require distinct mechanisms. Alternatively, however, the tight coordination of parsing and generation as defined in the *Dynamic Syntax* model of dialogue (Purver et al. (2006)) enables a straightforward account of how the context-dependence of both tasks allows participants to economise on processing.

3 Dynamic Syntax: A Sketch

Dynamic Syntax (*DS*) is a parsing-based approach to linguistic modelling, involving strictly sequential interpretation of linguistic strings. The model is implemented via goal-directed growth of tree structures and their decorations formalised using *LOFT* (Blackburn and Meyer-Viol (1994)), with modal operators $\langle\uparrow\rangle, \langle\downarrow\rangle$ to define concepts of *mother* and *daughter*, and their iterated counterparts, $\langle\uparrow_*\rangle, \langle\downarrow_*\rangle$, to define the notions *be dominated by* and *dominate*. *Underspecification* and update are core aspects of the grammar itself and involve strictly monotonic information growth for any dimension of tree structures and decorations. Underspecification is employed at all levels of tree relations (mother, daughter etc.), as well as formulae and type values, each having an associated *requirement* that drives the goal-directed process of update. For example, an underspecified subject node of a tree may have a requirement expressed in DS with the node decoration $?Ty(e)$, for which the only legitimate updates are logical expressions of individual type ($Ty(e)$); but requirements may also take a modal form, e.g. $?\langle\uparrow\rangle Ty(e \rightarrow t)$, a restriction that the mother node be decorated with a formula of predicate type. Requirements are essential to the dynamics informing the DS account: all requirements must be satisfied if the construction process is to lead to a successful outcome.

Structure is built from lexical and general computational actions. *Computational actions* govern general tree constructional processes, such as introducing and updating structure, as well as compiling interpretation for all non-terminal nodes in the tree, once individual leaf nodes are successfully decorated (with no outstanding requirements). This may include the construction of only weakly specified tree relations, characterised only as dominated by some node from which they are constructed (*unfixed nodes*), with subsequent update. Individual lexical items also provide procedures for building structure in the form of *lexical actions*, which are expressed in exactly the same terms as the more general processes, inducing both nodes and decorations. Thus *partial trees* grow incrementally driven by procedures associated with particular words as they are encountered, with a *pointer*, $\Diamond$, recording the parser's progress.

Complete individual trees are taken to correspond to predicate-argument structures. More complex structures can be obtained via a general tree adjunction operation defined to license the construction of a tree sharing some term with another newly constructed tree, yielding so-called *Linked trees* (Kempson et al. 2001). The resulting combined information from the adjoined trees is modelled as a conjunction of terms at the node *from* which the link is made. Importantly, adjunction, as other forms of construction and update, can be employed to model how subsequent speakers may dynamically provide fragmentary extensions in response to the previous utterance.

Structural as well as content underspecification play important roles in facilitating success-

ful linguistic interaction. The content underspecification of pronouns is represented as a place-holding metavariable, noted as e.g. $\mathbf{U}$, plus an associated requirement for update by an appropriate term value: $?\exists \mathbf{x}.Fo(\mathbf{x})$. Similarly, *names* are represented as initially introducing place-holders associated with a constraint providing the name of the individual entity picked out. For example, the name *Bill* contributes the decoration $\mathbf{U}_{\mathbf{Bill'(U)}}, \mathbf{Ty(e)}$. The subscript specification is shorthand for a transition across a LINK relation to a tree whose top node is decorated with a formula $Bill'(\mathbf{U})$, the name being taken as a predicate or name specification of individuals thus restricting possible updates to the metavariable[2]. Names can thus be seen as a procedure for identifying the individual being talked about, with a logical constant (e.g. $m21$, $m23$ etc.) picking out uniquely this individual eventually replacing the metavariable on the emergent tree. According to the DS account, the update of metavariables can be accomplished if the context contains an appropriate term for substitution. *Context* in DS involves storage of *parse states*, i.e., the storing of partial tree, word sequence to date, plus the actions used in building up the partial tree.

A major aspect of the DS dialogue model is that both *generation* and *parsing* are goal-directed, with parsing as the underlying mechanism and generation parasitic on it. A hearer builds a succession of partial parse trees in order to achieve an interpretation of the speaker's message. A speaker is modelled in DS as doing exactly the same only (s)he also has available a *goal tree* representing what they wish to say. Each possible step in generation —an utterance of a word— is governed by whatever step is licensed by the parsing formalism, constrained via the assumed *subsumption* relation between the thus-far constructed "parse" (partial) tree and the goal tree. By updating their growing "parse" tree relative to the goal tree (via a combination of incremental parsing and lexical search), speakers produce the associated natural language string.

The DS model of dialogue requires defining and taking into account both the speaker's goal and parse trees, as well as the hearer's parse tree. For fragment construal, we are interested in the extent to which B has successfully parsed what A has said, with the ability at any stage to interrupt to ask for clarification, reformulate, or provide a correction, by either repeating the expression or producing an alternative. As we shall see, B's parse tree makes explicit where need of clarification or miscommunication occurs, as it will be at whatever node from which a sub-routine extending that node takes place. According to the DS model of generation, repeating or correcting a constituent of A's utterance is licensed only if B's goal tree matches or extends a parse tree updated with the relevant subpart of A's utterance. Indeed, this update is what B is seeking to clarify, correct or acknowledge.

Notice that because of the incremental definition of DS, B can reuse the already constructed (partial) parse tree in their context, thereby starting at this point, rather than having to rebuild an entire propositional tree or subtree (e.g. of type e). Exploiting the assumed parity of representations in this way enables hearers to provide immediate feedback to the previous speaker, the effect being to narrow the focus on particular aspects of the interpretive space. The advantage of this emerges in the unified characterisation of any type of *ellipsis* construal as strictly context-dependence. Since context in DS involves the storing of current partial tree, word sequence to date, plus the actions used to date to build the partial tree, ellipsis construal can target any of those stored elements. In particular, for split/joint utterances, this enables switch from hearer to speaker at any arbitrary point in the dialogue, without such fragments having to be interpreted as propositional in type. This can then capture the dynamics involved in taking what the other speaker has just uttered, with the potential at any point to update it to accord with one's own emerging understanding of the interaction. In this way, speakers are able to guide each other's interpretations, and thus *jointly* narrow down as early as possible the burgeoning interpretive space.

4 NSU fragments in Dynamic Syntax

4.1 Non-repetitive Acknowledgement

From a DS perspective, phenomena like *reformulations* as in (1), or *extensions* to what one understands of the other speaker's utterance, (2), can be handled with exactly the same mechanisms as the sentence-internal phenomenon independently identifiable as *apposition* and illustrated below:

(5) A friend of my mother's, someone very

[2]These *linked* structures are suppressed in all diagrams.

famous, is coming to stay.

(6) Bob, the friend of Ruth's, is coming to stay.

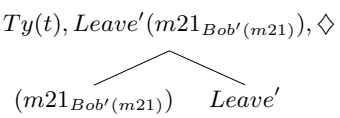
According to Cann et al. (2005), such structures are analysed as involving the building of paired terms, across a tree transition: the building of so-called *linked* structures relative to the constraint on such structures that they share a term. Reflecting this constraint, the update rule for such structures then takes the pair of type e terms so formed and licenses the building of a term whose compound restrictor is made up of the predicative content from each.

We now have the basis for analysing extensions and non-repetitive acknowledgements which build on what has been previously said by way of confirming the previous utterance. Recall examples (1) and (2). There are two ways for the processing of fragments which reformulate an interlocutor A's utterance: either (a) as interruptions of her, A's, utterance in which case immediate confirmation of identification of the individual concerned is provided, see (2), or (b) as confirmations/extensions of A's utterance after the whole of her utterance has been integrated, see (1). Both are modelled by DS.

Turning to (1), B's response *(Yeah,) the accounts guy* constitutes both a reformulation of A's utterance, as well as an extension of A's referring expression, in effect providing the appositive expression 'Bob, the accounts guy'. This means that B has processed A's original utterance, according to some identification of the individual associated with the name *Bob*: that is to say, they have constructed a full content representation for this utterance. B's reformulation has the effect of acknowledgement because it signals to A that he has processed/understood her asserted content, and, moreover, has no objection to the content, *unless* mistaken in that identification.

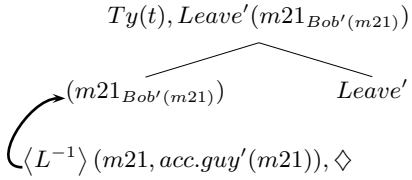
In DS terms, B's context consists of the following tree after processing A's utterance:

(7) B's Context for 'Yeah'



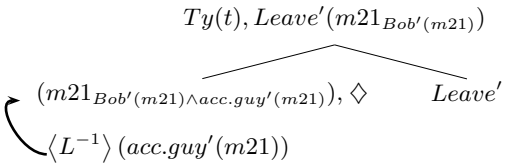
It is now open to B to re-use this representation, stored in his context, as the point of departure for generating the expression *the accounts guy*. In this case his own goal tree will now be decorated with a composite term made up both from the term recovered from parsing A's utterance and the new addition. Simplistically, all this requires is attaching a *linked* tree to the correct node, and then processing the content of the apposition in order to produce the words required. The defined steps include shifting the pointer to the appropriate node, projection of a *linked* tree from that node and processing the words *the accounts guy* (the *linked* tree is condensed below):

(8) B's "parse" tree licensing production of *the accounts guy*: LINK adjunction



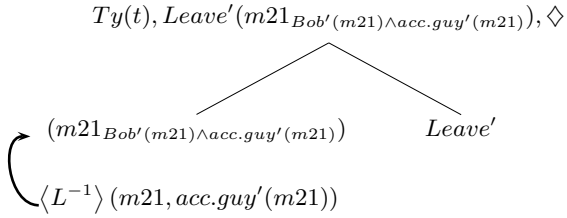
Further updating this representation according to the DS processing protocol requires adding the acquired restrictions at the node from which the *linked* tree is projected:

(9) Updating B's "parse" tree licensing production of *the accounts guy*



For the sake of brevity, the full sequence of parsing steps leading to this result are suppressed in what follows. It essentially involves processing the expanded structure for the term currently decorating the source node, $(m21_{Bob'(m21)})$, and then extending this in turn with the additional structural representation for *the accounts guy*, i.e. $(acc.guy'(m21))$. Finally, the information is passed up to the top node of the main tree, completing the parse tree to match B's goal tree in uttering the expression *the accounts guy* :

(10) Completing B's "parse" tree licensing production of *the accounts guy*



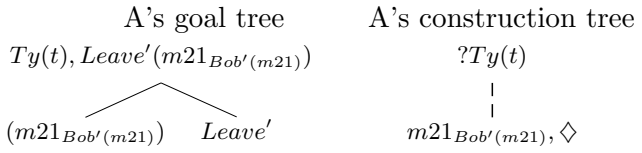


Figure 1: Licensing production of a correction by *ADJUNCTION

4.2 Non-repetitive Clarification

In the acknowledgement case above, the term relative to which the *linked* structure is built is fixed; but the very same mechanism can be used when the interlocutor needs clarification. In (2), B again takes as his goal tree a tree decorated with an expansion of the term constructed from parsing A's utterance but nevertheless picking out the same individual. Using the very same mechanism as in (1) of building a *linked* structure constrained to induce shared terms, B provides a distinct expression, the name *Chorlton*, this time before he has completed the parse tree for A's utterance. This name, contributing a metavariable plus the constraint that the individual picked out must be named *Chorlton*, is used to decorate the linked node so that it makes explicit the additional predicative constraint on the individual being described. The outcome of this process, when the linked structure is evaluated, is a composite term $m21_{Doctor'(m21) \wedge Chorlton'(m21)}$. This process, that is, is identical to that employed in B's utterance in (1), though to rather different effect at this intermediate stage in the interpretation process. This extension of the term is confirmed by A, this time trivially replicating the composite term which processing B's utterance has led to (see Kempson et al 2007 for such trivial goal tree-parse tree matches). The eventual effect of the process of inducing *linked* structures to be decorated by coreferential type e terms may thus vary across monologue and different dialogue applications, but the mechanism is the same.

4.3 Correction

It might be argued nonetheless that correction is intrinsically a dialogue phenomenon. In (4) for example, reproduced below:

(4) A: Bob left.
B: Rob?
A: (No,) (Bob,) the accounts guy.

B has misheard and requests confirmation of what he has perceived A as saying. A in turn rejects B's utterance and corrects it. Presuming rejection as simple disagreement (i.e. the utterance has been understood, but judged as incorrect), in DS terms, this means that A has in mind a goal tree that licensed what she had produced, which is distinct from the one derived by processing B's clarification. As shown in Kempson et al. (2007), this means that A has been unable to process B's clarification request as an extension of her own context. Instead she can parse the clarification by exploiting the potential for introducing an initially structurally underspecified tree-node to accommodate the contribution of the word *Rob*. Subsequently by utilising the actions stored in context previously by processing her own utterance of the word *left* she is able to complete the integration of the fragment in a new propositional structure.

In order to produce the following correction, what is required is for A to establish as the current most recent representation in context her original goal tree. This can be monotonically achieved by recovering and copying this original goal tree to serve as the current most immediate context[3]. Under these circumstances, given the DS grammar-as-parser perspective, several strategies are now available. A is licensed to repeat the name *Bob* by locally extending the node in the context tree where the representation of the individual referred to is located by using the rule of LATE*ADJUNCTION, a process which involves building a node of type e from a dominating node of that type (illustrated in Kempson et al. 2007). An alternative way of licensing repetition of the word *Bob* is to employ one of the strategies generally available for the parsing of long distance dependencies i.e. constructing initial tree nodes as unfixed (*ADJUNCTION). Starting with Fig 1 above, illustrating the introduction of the unfixed node, we show here how the latter strategy can be exploited to license the production of the fragment.

[3]Corrected representations must be maintained in the context as they can provide antecedents for subsequent anaphoric expressions.

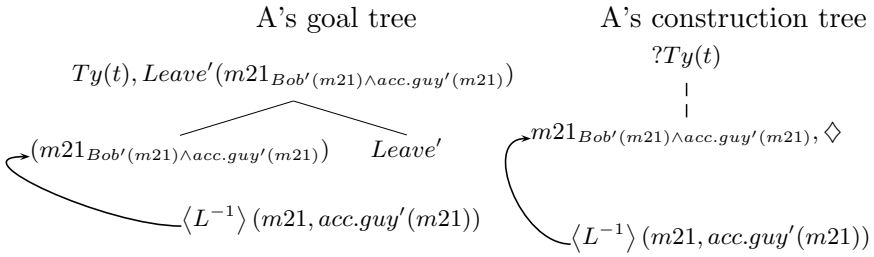


Figure 2: Licensing the production of correction and extension (1): LINK ADJUNCTION

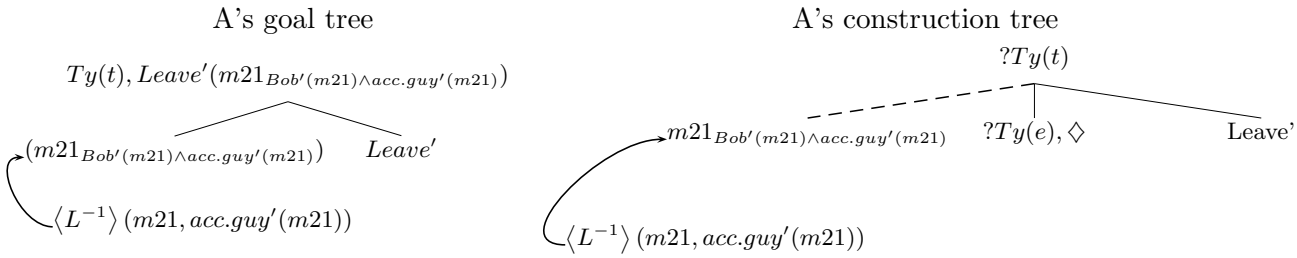


Figure 3: Licensing the production of correction and extension (2): retrieving and rerunning the actions for *left*, pointer return to subject node

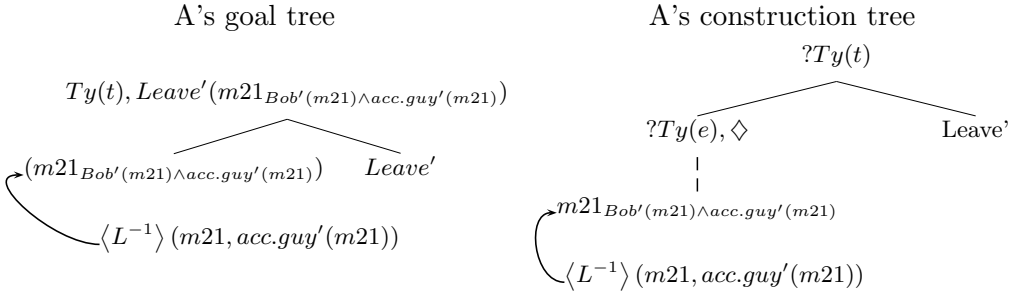


Figure 4: Licensing the production of correction and extension (3): preparation for UNIFICATION

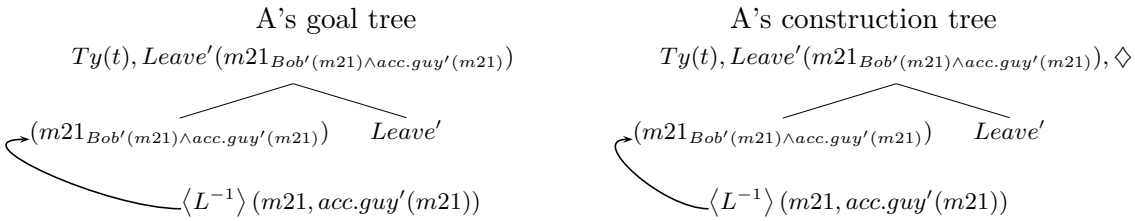


Figure 5: Licensing the production of correction and extension: completed tree matching the goal tree

An option available to A at this point is to introduce, in addition or exclusively, a reformulation of her original utterance in order to facilitate identification of the named individual which proved problematic for B previously. She can answer B's utterance of *Rob* with *(No,) (Bob,) the accounts guy*, as in (4) or simply with *(No,) the accounts guy*. Both are licensed by the DS parsing mechamism without more ado. The structure derived by processing such an extension is exactly that of (1) above (compare goal tree in Fig 2 above and tree in (10)). As we mentioned before, *context*, as defined in DS, keeps track not only of tree representations and words but also of the actions contributed by the words and utilised in building up the tree representations. Production of the correction in (4) is licensed to be fragmental because the original actions for parsing/producing the word *left* are available in the context and can be recalled to complete the structure initiated by processing/producing the name *Bob* (see Fig 3-5 above for the licensed parsing steps).

4.4 Structure and Dialogue Function

In the examples considered so far, we have seen how a single type of mechanism can serve distinct functions. A more striking case is (3), where the hearer, B, is able to leap to a hypothesis as to how A's question is going to be completed, and provides that completion by way of answer. Here we have the case where more than one function can be fulfilled even by a single utterance. As in (1)-(2), license for such a use turns on taking the context that was constructed by parsing input from the interlocutor as the point of departure. That B is extending the structure set up by A's utterance is self-evident; but in addition, both A's utterance, if she had completed it, and B's utterance, as presented, are elliptical as to the second disjunct. The success of this particular form of split utterance turns on the fact that what A is presenting is a duplex *yes-no* question with both possible answers provided by the two disjuncts. So in completing it by providing just the second disjunct, B can succeed in answering the question while simultaneously completing it. Though there is more to say about the distinctive properties of *or*, the significance of (3) here lies in the use of the single expression *right-handed* to fulfil two functions, both the completion of a question and the provision of an answer. In DS this can be modelled, reflecting the phenomenon itself, without having to assume the superimposition of two distinct structures, one upon the other. Incidentally, this is a case contradicting what is supposedly unique to such interrupting completions, namely, that they require acknowledgement by the hearer before proceeding.

5 Conclusion

As these fragments and their construal have demonstrated, despite serving distinct functions in dialogue, the mechanisms which make such diversity possible are general strategies for tree growth. In all cases, the advantage which use of fragments provides is a "least effort" means of re-employing previous content/structure/actions which constitute the *context*. As modelled in DS, it is more economical to reuse information from context rather than constructing representations afresh (via costly processes of lexical retrieval, choice of alternative parsing strategies, etc.).

A further quandary in dialogue construal is that, despite such avenues for economising their efforts, interlocutors are nevertheless faced with an increasing set of interpretative options at any point during the construction of representations. One option available to hearers is to delay a disambiguating move until further input potentially resolves the uncertainty. However, as further input is processed and parsing/interpretive options increase potentially rapidly, maintenance of these open options becomes difficult for a human processor. The incremental definition of the DS formalism allows for the modelling of an alternative available to hearers: at any point they could opt to intervene immediately, and make a direct appeal to the speaker for more information at the maximally relevant point during construction. It seems clear that the latter would be the favoured option and this is what clause-medial fragment interruptions as in (2) illustrate.

The phenomena examined here are also cases where speaker's and hearer's representations, despite attempts at coordination may, nevertheless, separate sufficiently for them to have to seek to explicitly "repair" the communication (see especially (4)). In the model presented here the dynamics of interaction allow fully incremental generation and integration of fragmental utterances so that interlocutors can be taken to constantly provide optimal evidence of each other's representations so that necessary adjustments can be timely introduced. Thus, in this model, there is no need for the employment of further structures/strategies specific to the particular dialogue function to which such fragments are put.

References

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

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

Raquel Fernández. *Non-Sentential Utterances in Dialogue: Classification, Resolution and Use*. PhD thesis, King's College London, University of London, 2006.

Ruth Kempson, Wilfried Meyer-Viol, and Dov Gabbay. *Dynamic Syntax: The Flow of Language Understanding*. Blackwell, 2001.

Ruth Kempson, Andrew Gargett, and Eleni Gregoromichelaki. Clarification requests: An incremental account. In *Proceedings of the 11th Workshop on the Semantics and Pragmatics of Dialogue (DECALOG)*, 2007.

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