Grammar Resources for Modelling Dialogue Dynamically

Ruth Kempson[†], Andrew Gargett[†], Eleni Gregoromichelaki[†],
Matthew Purver[‡], Yo Sato[§]
[†]King's College London, Strand, London WC2R 2LS, UK
{ruth.kempson,andrew.gargett,eleni.gregor}@kcl.ac.uk
[‡]Queen Mary University of London, Mile End Road, London E1 4NS, UK
mpurver@dcs.qmul.ac.uk
[§]University of Hertfordshire, College Lane, Hatfield AL10 9AB, UK
y.sato@herts.ac.uk

Abstract

This paper argues that by analysing language as a mechanism for growth of information (**??**), not only does a unitary basis for ellipsis become possible, otherwise thought to be irredeemably heterogeneous, but also a whole range of sub-types of ellipsis otherwise thought to be unique to dialogue. Dialogue fragment types modelled include *reformulations*, *clarification requests*, *extensions*, and *acknowledgements*. Buttressing this analysis we show how incremental use of fragments serves to incrementally narrow down the otherwise mushrooming interpretational alternatives in language use, and hence is central to fluent conversational interaction. We conclude that, by its ability to reflect dialogue dynamics as a core phenomenon of language use, a grammar with inbuilt parsing dynamics opens up the potential for analysing language as a mechanism for communicative interaction.

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) conversation demonstrates genre-specific characteristics, for which a grammar specific to such activity must be provided (**???**); or (b) the cross-speaker flexibility and variation characteristic of dialogue has to be seen as due to the specifics of the parsing/production systems which are based upon, but nevertheless distinct from, the mode-neutral grammar characterising the individual's competence in that language. Both alternatives raise issues for an empirical account of what language users' knowledge of a given language amounts to. One core phenomenon where such issues are vividly displayed is *ellipsis*, displayed both in monologue and dialogue genres. In simultaneously being the sine-qua-non characteristic of dialogue while nevertheless being subject to restrictions diagnostic of grammar-internal mechanisms, ellipsis would seem to straddle the remits of dialogue-modelling

and grammar-modelling in a way that does not allow for a unified explanation given conventional grammar formalisms. In order to preserve the separation between discourse phenomena and core grammar constraints, this has been taken to indicate a division between discourse ellipsis and grammar-internal ellipsis (**?**), confirming what is widely taken to constitute the heterogeneity of ellipsis phenomena (**?**). But such a move notably fails to reflect that the single most defining feature of all types of ellipsis is systemic context-dependence. In turn, such an apparent split between discourse-based and grammar-constrained ellipsis has been taken to vindicate the claim that a dialogue-specific grammar is necessary to deal with dialogue elliptical phenomena.

In this paper, to the contrary, we will show that ellipsis phenomena in general are characterisable by grammar-internal mechanisms without any stipulation specific to their use in dialogue. This result is achieved by developing a grammar formalism that directly reflects the dynamics of language processing as the basis for explaining structural properties of language. More specifically, the grammar is action-based and tracks the incremental and progressive growth of interpretation in real time. In this respect, the methodology of preserving a gulf between the design of the grammar-formalism and performance considerations is no longer sustained. The concept of *procedures*, equivalently *actions* determining growth of represented information, is central to the formal specification of the grammar itself, not a characteristic of just the parser or the generator.

The approach within which we set out our account of fragment construal, is that of *Dynamic Syntax* (*DS*) (**??**). In this framework, the usually static notion of "syntax" is replaced by "syntax" as the progressive construction of semantic representations set in context, following the dynamics of parsing. Language production is presumed to be parasitic on the very same tree-growth processes, following the dictates of the grammar formalism. The only difference between the tasks is that, in production, the speaker has some concept of what it is that they are saying against which the application of the tree-construction rules has to be checked.

Such a framework is well suited to modelling ellipsis in all its forms: what is more standardly seen as diverse grammar-internal constraints on ellipsis are handled in DS as falling out from the definition of the dynamics of the time-linear growth of the semantic representation. This growth is modelled as consisting of three dimensions: (a) growth of the string consumed by the parser, i.e. the set of words processed increases; (b) growth of the semantic tree-representation, i.e. the trees are updated and become more specified; and (c) the set of actions used in constructing tree representations accumulates. In **?** and **?** it is argued that *context* should then be accordingly defined as incorporating a record of words processed, structures established, and the actions that led to this structure. The latter is what allows DS to deal straightforwardly and in a unified way with cases of ellipsis. In particular, both semantic and syntactic restrictions on fragments are dealt with in the same way, as constraints on tree growth. Sets of actions/procedures, what lexical items are presumed to consist in, are sequentially executed and subsequently stored in the context record. They can then be available for recovery and re-use, and some types of ellipsis crucially rely on this feature of context for their resolution. So, on this account, the concept of *procedure* is central to context too.

In this paper we adopt this analysis as background for demonstrating how a range of elliptical phenomena which might at first sight seem specific to conversational dialogue – acknowledgements, clarifications, reformulations, utterance-exchanges involving interruption and intra-sentential switch of speaker-hearer roles – can be analysed without any stipulation specific to such functions. The immediate advantage of such an account is that it preserves, indeed directly reflects, the intuition that ellipsis occurs when the context fully determines construal as far as the message conveyed is concerned except for the elliptical fragment presented. We shall conclude that grammars that are defined as dynamical systems directly reflecting mechanisms for growth of interpretation in real time also allow for dynamic adjustment of the information conveyed according to the feedback received by the other interlocutor. Accordingly, grammars can be seen as a mechanism for communicative interaction: the human capacity for language is thus grounded directly in the interactional activity which it serves.

2 Background

In natural language use, people often talk to each other with apparently fragmentary utterances, a phenomenon broadly known as ellipsis:

(1) A: Have you seen Mary?
B: (a) Mary?
(b) No, I haven't.
(c) But I have Bill.
(d) Tom too.
(e) Could either of them help?

Despite a robust folk intuition that ellipsis can occur whenever the context makes obvious how the apparently fragmentary expression of thought is to be "completed", ellipsis has very generally been seen as a heterogeneous set of phenomena not subject to a uniform explanation. The folk intuition is thus simply set aside without discussion: indeed it does not play any role in orthodox accounts of ellipsis.

Amongst the rich array of ellipsis effects, there are cases of elliptical fragments where the the linguistic surface form of the antecedent provides the resolution. These cases have been subject to a great deal of study over recent years - *VP-ellipsis* as in (1b), *sluicing*, *stripping* (as in (1d)), *gapping*, *pseudo-gapping* (as in (1c)), etc. Each sub-type is said to display different structural constraints on their construal, justifying their distinct analyses. What brings these together is that they, in some sense, constitute a complete sentence, given their interpretation as suitably completed by the provided antecedent, so that ellipsis can be seen as requiring syntactic/semantic rules defined over conjoined sentences (though cf. **?**).

But, even for these cases, ellipsis remains very puzzling, being seen as unresolvably heterogeneous. There are cases where relative to a single antecedent source, ambiguity may arise, with a single string admitting either a strict or a sloppy reading:

(2) John checked over his mistakes, and so did Bill/Bill too.
'Bill checked Bill's mistakes' (sloppy)
'Bill checked John's mistakes' (strict)

This phenomenon would seem to be broadly a phenomenon of construal, requiring a semantic basis for explanation under which the CONTENT of the antecedent has somehow to be massaged by making available alternative forms of abstraction to match what is required at the ellipsis site (hence the classic account in terms of various abstraction operations applicable to the content of the first conjunct to yield a novel predicate applicable to the subject of the second (**?**). However, elliptical phenomena also appear to be sensitive to the very same constraints that affect surface syntactic operations. For example, relative clause constructions interfere with "movement" operations (they are *islands*). The same interference seems to be responsible for the ungrammaticality of the VP-ellipsis construction in (4) below:

(3) John interviewed every student who Bill already had.

(4) *John interviewed every student who Bill ignored the teacher who already had.

What is displayed in (3) is the phenomenon of *antecedent contained deletion* (or *antecedent-contained ellipsis*), so called because the ellipsis site appears to be contained within the antecedent from which its own interpretation is built up. Circularity apart (for which different solutions are promulgated depending on the framework), the problem is that such structures preclude any binding by the relative pronoun, *who*, across a relative-clause boundary. Accordingly, (4) is ungrammatical because there is no possibility of resolving the ellipsis site indicated by *had* by means of the containing matrix verb in the way that (3) appears to allow. This pattern is reminiscent of restrictions on so-called long-distance dependencies which are taken to be diagnostic of a syntactic process:

(5) The man who Sue is worried that her sister is planning to marry lives in Austria.

(6) *The man who Sue is concerned about her sister who is planning to marry lives in Austria.

Though (5) is wellformed with *who* construed as the object of *marry*, no dependency for the first relative pronoun in (6) is possible, because there is a further relative clause boundary (an island) between the relative pronoun *who* and the same verb *marry* that *who* has to be associated with (the observation goes back to (**?**)). Because such a restriction is not expressible in semantic terms – for example, the lambda calculus (the logic taken to underpin semantic combinatorics) would impose no such restriction – they have been taken as evidence for a concept of syntax independent of both semantics and phonology, and a diagnostic of what constitutes a syntactic process. Some types of ellipsis at least are thus argued to be within

the remit of natural-language syntax, involving low-level deletion of phonological material (PF Deletion: cf **???**).[1]

2.1 Dialogue Ellipsis

What all these grammar-internal characterisations of ellipsis miss, given their remit of characterising only sentence-internal properties, is the broad array of elliptical effects in dialogue. One of the most striking characteristics of conversational dialogue is the extent, and freedom, with which participants make use of utterance fragments. Indeed fragmentary expressions that occur in dialogue apparently allow interpretations that indicate many sorts of conversational interaction. Interlocutors can extend each other's utterances, while at the same time displaying their acceptance/understanding of the other's presentation:

(7) A: Bob left.
B: the accounts guy, (yeah).

They can interrupt and finish each other's utterances:

(8) Conversation from A and B, to C:
A: We're going to
B: Bristol, where Jo lives.

They can even use each other's utterances as the basis for what they themselves have to say, without waiting for their interlocutor to finish:

(9) A: Most of the ones that we brought seem to have erm
B: survived
B: survived. Which I'm glad. [from (**?**)]

And, as these examples illustrate, such switch of roles between hearer and speaker can take place across any syntactic dependency whatsoever: across a determiner-noun dependency (11), across a preposition-NP dependency (8), or auxiliary-verb dependency (9).

Such fragmentary utterances may be interpretable only relative to partial contents currently being presented by other interlocutors. Such utterances seem in some sense to be constructed jointly by participants, relying on feedback by clarification, disagreement, or correction (henceforth, A female, B male):

(10) A: Have you mended
B: any of your chairs? Not yet.

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

[1] Further cases indicate ellipsis construal sensitivity to morphological idiosyncracies (see for discussion **????**).

(12) 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 slide [unclear] on my heart. [BNC: KPY 1005-1008]

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

It might seem, and it has traditionally been assumed, that the grammar/syntax, which under orthodox assumptions has to do with sentence strings and propositions, should have nothing to say about such fragmented dialogue turns. But firstly, as was pointed out early on (**?**), there are grammatical sequences of words other than sentences, as can be seen in all the examples above, and the grammar should be able to characterise those and distinguish them from plain "word-salad". Secondly, the grammar is responsible for characterising certain dependencies among lexical items in grammatical sentences. Consider then the licensing of the negative polarity item (NPI) in (10): Such NPIs are only licensed by appearing in sentences that contain some explicit "affective" element, namely negation, question etc. (REFS). Now, in (10), the only element that can license the NPI is the interrogative morphology registered at A's turn. It would seem then that A and B's turns should somehow be joined together to form a single utterance, otherwise we would not be able to match the intuition that the discourse is perfectly well-formed. Now one might be tempted to conclude that, indeed, this vindicates a grammar which characterises sentence strings as it would seem that this licensing occurs only when we assume that a single string of words is spread over two turns. But this would be too hasty: Notice what happens with the second A and B exchange in (11). The licensing of the reflexive anaphor *myself* is only possible because its antecedent, namely B, is part of the CONTENT of the turn started by A. But if we now try to join the two strings together (*but did you burn myself*) the result is not what the exchange was meant to convey. From that point of view, it seems that this phenomenon of turn-sharing, loosely characterised as *split utterances*, has to do with the sharing of contents rather than strings.

Given this conclusion, one might consider that what is needed is a grammar for dialogue, with the specific remit of determining how content is jointly constructed over turns by interlocutors. Since content computation is known to rely crucially on context, such a grammar would have to regiment the pragmatic inferencing that results in the derivation of what interlocutors actually do with their utterances, including specification of the utterance's speech-act function in the dialogue. CHECK WHETHER THIS IS WHAT THE RHETORIC WAS SUPPOSED TO ADDRESS Let's see whether this would be a desirable move according to the data presented here. While the NP fragments in (12)–(13) might be characterised as distinct utterance types, serving rather different functions of clarification, acknowledgement, correction, they also illustrate how speakers and hearers may contribute to the joint enterprise of establishing some shared communicative content. In such examples, each speaker contributes parts to a single collaborative

utterance, thus these also fall under the phenomenon loosely characterised as *split utterances*, despite the fact that both speakers may well not have the same content in mind guiding what they say, nor is their contribution intended to have the same function. Even (7), an *acknowledgement*, can be seen as part of a split utterance, for it is similar in form to an afterthought *extension* added to A's sentential utterance. As (12) shows, joint construction of content can proceed incrementally: B provides a *reformulation* as a *clarification request*, resolved by A within the construction of a single proposition. In (13), the fragment A provides functions as a *correction* of some aspect of B's understanding, with A and B having to negotiate as to whose information is more reliable in order to secure coordination. Nevertheless the correction in (12) also constitutes an extension, so that a single conjoined propositional content is derived DURING which coordination is achieved.

Such diversity of fragment uses might, then, seem evidence of conversation-specific rules as part of a grammar. Taking such a line, **?** presents a thorough taxonomy, as well as detailed formal and computational modelling of *Non-Sentential Utterances*, referring to contributions like (7) as *repeated acknowledgements* involving *reformulation*, with type-specific "accommodation rules". **?** 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*. In this respect, any utterance function that can be accomplished by a single NP, e.g. *the accounts guy*, has to be characterised by postulating appropriate grammatical and contextual rules that become part of the dialogue management process. If what speakers and hearers can perform and understand in dialogue is limitless and non-deterministic, shaped through feedback, this might not be a desirable move. And what such an account fails to bring out is how these phenomena are structurally replicated in monologue in apposition structures, possibly extraposed:

(14) Bob, the accountant, is coming to stay.

(15) Bob left, the accountant.

Furthermore, even these construction-specific analyses of fragments in dialogue are taken to require a sentential form of analysis: the fragment is assigned a type which in combination with a suitable abstract with respect to what is provided in context will yield a sentential form of meaning. But, as (13) indicates, such fragments can be used at early points in a dialogue exchange when there may be no context-provided content relative to which the fragment can be provided a suitable abstract; but nonetheless the participants in the dialogue can exchange a clarificatory request and reply so that the communicative exchange can proceed.

The challenge posed by split exchanges for orthodox grammar-parser/production mechanisms is considerable, and they have not been much addressed (though see **?**). The problem is as follows: speakers shift into hearing as though they had been hearing all along: parsers shift into production as though they had been speaking all along. But, until recently, parsing and generation systems have been defined

relative to a grammar whose remit is complete sentence-strings. And, even now, though parsing and generation systems are increasingly reflecting incrementality (**???**), such incrementality must generally come from the processing model, with the grammar defined statically and independently. Yet, to deal with split utterances, parsing/generation systems have to be defined with a flexibility allowing either one to take up from where there has been a switch, despite the fact that both string preceding or following the switch may fall outside the set of strings licensed as well-formed by the grammar. This problem is not one that confronts a grammar defined in terms of incremental growth of interpretation and, as a corollary, context. To the contrary, if by assumption the same mechanisms for monotonic tree-growth are shared by both systems, it is this type of framework that, uniquely, can model this unproblematic shift of role as a wholly natural, indeed a predictable consequence. This suggests that incrementality in both generation and parsing, and the potential to provide update from whatever intermediate point speaker-hearer switch takes place, is at the core of the explanation for the prevalence of split utterances in conversation.

Accordingly, we now turn to exploring the use of grammar-internal resources to capture such phenomena. In this type of grammar, as we shall see, it will be actions/procedures for interpretation which replace the static notions of (representations of) denotational content itself as the central notion.

3 Dynamic Syntax: A Sketch

The conceptual starting point is that conversational dialogues emerge incrementally over the course of the interaction through the distinct contributions of the participants, each employing no more than resources internal to the grammar.

3.1 Introduction

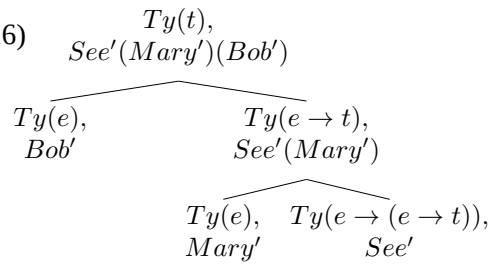
Dynamic Syntax (*DS*) has three main characteristics underlying the modelling of how contextual information can be incorporated AS IT ARISES with linguistic information during interaction in dialogue. First, DS provides a fully incremental parsing model, based on a strictly monotonic process of interpretation update. process of update is goal-directed as it is dynamically driven by *requirements* for update at each initial or intermediate stage. As will be shown shortly, these two characteristics are closely related, and together allow for the third DS distinguishing feature, the fact that the same mechanisms are exploited to model both parsing (= understanding) and generation (= production).

3.2 Content representations

Taking up the DS formulation of context first and its tree-theoretic representation, content is modelled as formulae annotating trees which are formalised using *LOFT* (**?**), a modal logic designed to allow the processor to refer to partial, complete or required tree structure. LOFT makes available modal operators $\langle\uparrow\rangle, \langle\downarrow\rangle$ defining the concepts of *mother* and *daughter* nodes and their iterated counterparts, $\langle\uparrow_*$

$\rangle, \langle\downarrow_*\rangle$, defining the notions *be dominated by* and *dominate* thus yielding partially specified trees in terms of structure. Annotations (*decorations*) on those nodes determine *formula*, *type* and *tree-node* position, and a pointer, $\Diamond$, indicates the node under development. Complete individual trees are taken to correspond to predicate-argument structures, with nodes in such trees decorated with sub-terms of the propositional formula decorating the root. Formulae are composed using the combinatorics of the lambda calculus (however we omit the lambda-binding operators in the graphics below to simplify illustration):

(16)



The representation language for content in DS is the *epsilon calculus*, a conservative extension of predicate logic. Thus quantificational NPs contribute terms of low type e rather than some higher generalised-quantifier type (**?**). Their processing results in the construction of the natural-deduction counterpart of a quantifier, i.e., terms in the epsilon calculus. For example, the natural language NP *a man* contributes the content $(\epsilon, x, Man'(x))$ which stands for the witness of the formula $\exists x Man'(x)$ according to the following equivalence:

(17) $\dfrac{\exists x \phi(x)}{\phi(\epsilon, x, \phi(x))}$ $\qquad$ $\dfrac{\text{predicate logic formula}}{\text{epsilon calculus equivalent}}$

The advantage of such terms is that they can be extended to include their context of occurrence inside their restrictor. So, for example, the semantic evaluation rules for a proposition derived from the string *A man cried* will produce not simply the term $\epsilon, x, (Man'x)$ but instead $\epsilon, x, (Man'x \wedge Cry'x)$ which denotes a witness of the set of men who cried:

(18) $Ty(t), Cry'(\epsilon, x, Man'x) \quad \mapsto$
$Ty(t), Cry'(a) \quad$ where $a = (\epsilon, x, (Man'x \wedge Cry'x))$

This allows a straightforward account of cross-sentential anaphoric dependence on quantifier antecedents (E-type anaphora: Evans 1980 and many others since), as the term abbreviated as a above can now serve as the referent associated with with *he* in a possible continuation *He was upset*.[2] Similarly, names and definites can be analysed as *iota terms* (epsilon terms of widest scope). So a name like *Bill* or a definite description like *the man* will contribute terms $(\iota, x, Bill'(x))$ and $(\iota, x, Man'(x))$

[2]Relative scope is expressed not on the tree itself, but involves incremental collection of scope-dependency constraints (either lexical or structurally determined) with the output formulae and such set of scope dependencies being subject to an evaluation algorithm determining their combined effect of interpretation.

to the tree representation (in the graphics below we omit the epsilon terms for simplicity of illustration unless they have a bearing on the analysis).[3]

3.3 Dynamics

DS employs an underspecification-plus-enrichment model of update in context. Underspecification is employed at all levels of tree relations (mother, daughter etc.), as well as formulae and type values, each with an associated *requirement* driving the goal-directed process of update. Tree updates are executed with lexical and general computational actions, which can be understood to be transition functions from one tree to another. *Computational actions* govern general tree-constructional processes, introducing/updating structure and 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 through *unification* with some other node (unlike **?**, partial trees are part of the model).

Individual lexical items also provide *lexical actions* for building structure, expressed in exactly the same terms as the more general processes, inducing nodes and decorations. Thus *partial trees* grow incrementally, driven by procedures associated with words as encountered, with the pointer, $\Diamond$, tracking the parse progress, and thus taking care of word-order – see Figure 1.

An expected tree starts out as a node with the decoration $?Ty(t)$, an entirely underspecified tree requiring a proposition, but this will be enriched or specified progressively. A node in general may, for example, be specified so as to determine that its only legitimate updates are logical expressions of individual type ($Ty(e)$), or the requirement may also take a modal form, e.g. $?\langle\uparrow\rangle Ty(e \rightarrow t)$, a restriction that the mother of this node be decorated with a formula of predicate type. Requirements are essential to the DS dynamics: all requirements must be satisfied if the construction process is to lead to a successful outcome.

3.4 Parsing vs. Generation

A crucial feature of the DS dialogue model is that it is *bi-directional*: parsing and generation use the same action definitions and build the same representations, following essentially the same procedure of left-to-right updates through actions. This is due to DS's monotonic incrementality and goal-directedness. In fact, we propose that these two features can be characterised as closely related through *predictivity*, a term often used in literature on human online processing (e.g. **?**). Generally incrementality is rather loosely characterised as information growth at each word input, but can be more precisely defined as the input word being incorporated into a predicted structure, in a manner close to the notion of *connectedness* in some psycholinguistic literature (**??**). Information growth, that is, is always ensured at

[3] The account of names and definites is simplified for exegesis, but see **?**.

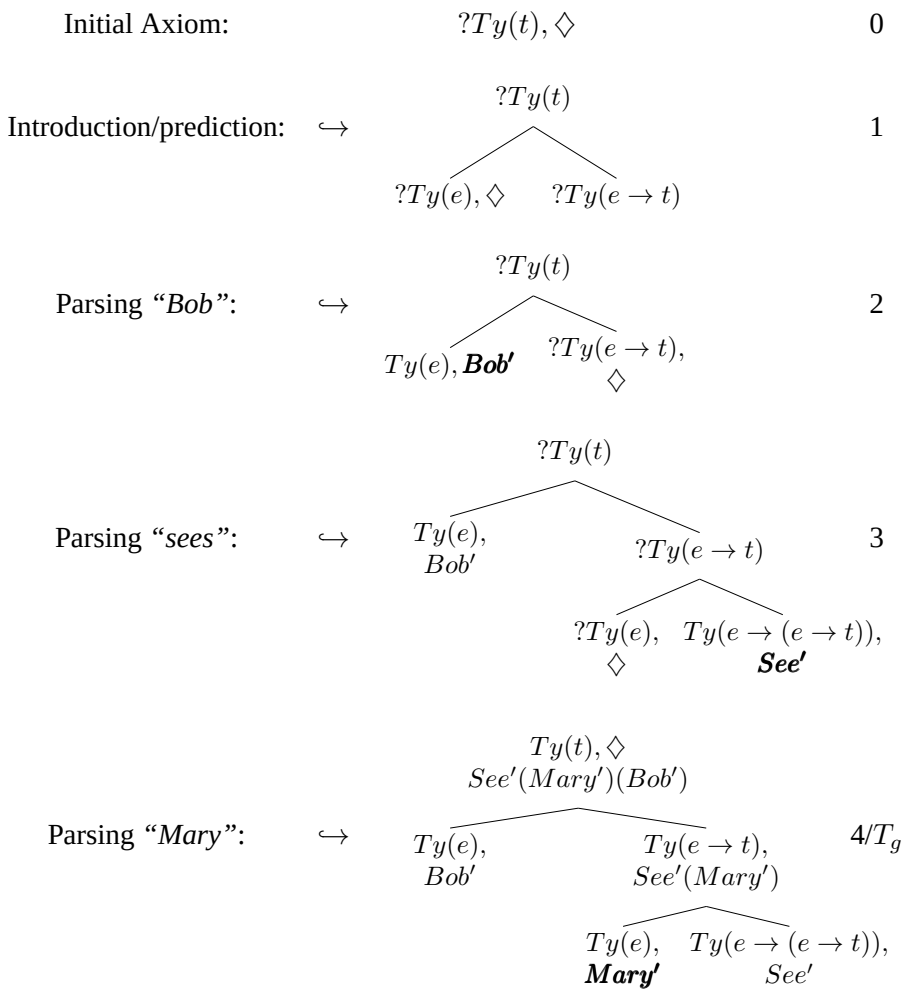


Figure 1: Monotonic tree growth in DS

each input, as a word (lexical action) either fulfills a requirement or creates a requirement as well as contributing the word meaning. Monotonicity is also ensured, as what is 'required' or predicted initially in a DS parse is the undespecified type t node, and a series of more specific trees are created from this as the parse proceeds.

It is this predictive nature that renders DS bi-directional. In parsing, the hearer builds a succession of partial parse trees, of course without record of what the eventual proposition is going to be but with that partial tree including predictions about what can follow, in the form of as yet unsatisfied type requirements (see Figure 1 steps 0-4). Generating the same sentence proceeds in exactly the same fashion, provided that a goal tree T_g (tree 4 in Figure 1) is available for the speaker representing what they wish to say. Each possible step in generation, the uttered word, each in turn, is governed by whatever step is licensed by the parsing formalism, as further constrained by the required *subsumption* relation of the thus-far constructed

"parse" (partial) tree to 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.

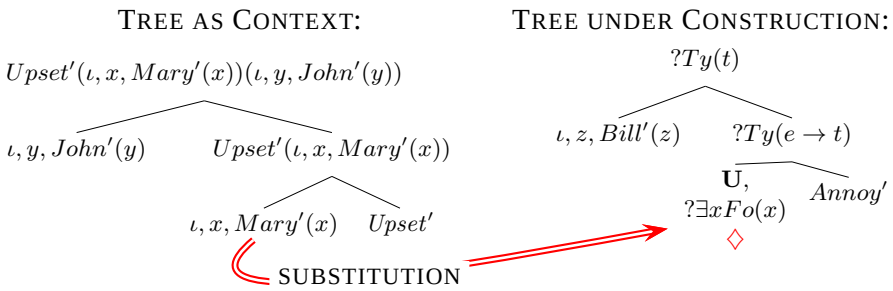
4 Context, anaphora and ellipsis in DS

4.1 Anaphora and relative clauses

Content and structural underspecification both play important roles in facilitating successful linguistic interaction. Linguistic items like pronouns are paradigm cases of such underspecification in terms of their content. This type of content underspecification is represented in DS as involving a place-holding metavariable, noted as e.g. **U**, **V** etc., plus an associated requirement for replacement by an appropriate term value: $?\exists \mathbf{x}.Fo(\mathbf{x})$. This value has to be supplied by the context of the discourse. *Context* in DS involves storage of entire parse states which includes a record of the words processed to date, the tree structures built up and the actions utilised to build these structures. So consider the parsing of B's utterance below in the context of A's utterance:

(19) A: John upsets Mary.
B: Bill annoys her.

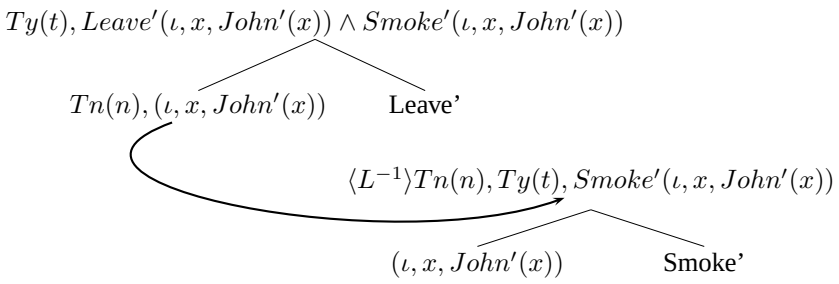
(20)



The process of *Substitution* licenses copying of a term in the context tree to replace a metavariable awaiting replacement.

This context is sometimes constructed while processing the utterance itself. One instance of this phenomenon is the case of *relative clauses* in English, which require more complex structures than the simple binary predicate-argument structures we have seen so far. These are obtained via a general tree adjunction operation defined to license the construction of a tree sharing some term with another newly constructed one, 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. In such constructions, the relative pronoun provides a copy of the head noun inside the *linked* tree and the content derived on this tree is incorporated in the main structure as an extension of the term appearing as argument there. In other words, one partial tree is used as context for the processing of another:

(21) John, who smokes, left:

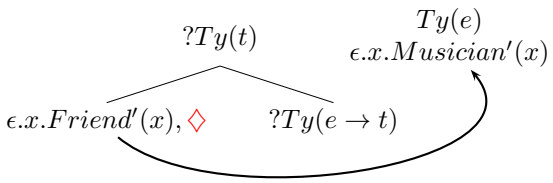


This concept of constructing trees in pairs is extendable to apposition devices. So, in (22)-(14), the constraint on linked structures as sharing a term is met through the construction of a compound term made up of a restrictor derived from the paired formulae (**?**):[4]

(22) A friend, a musician, smokes.

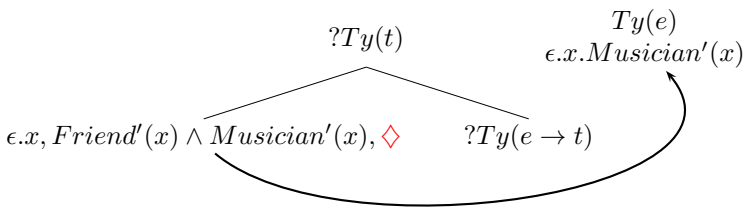
The construction step involves building a transition from a node of type e to a linked tree, also constructed to be decorated by a term of the same type:

(23) Parsing *A friend, a musician*



Evaluation of the *linked* nodes, both of type e, yields the composite term: $\epsilon.x, Friend'(x) \wedge Musician'(x)$:

(24) Parsing *A friend, a musician*



The final formula that is derived from the parse of *A friend, a musician smokes* is: $Smoke'(\epsilon, x, Friend'(x) \wedge Musician'(x))$.

4.2 Ellipsis

These various mechanisms are brought together in **???** to provide a uniform basis for ellipsis construal. They argue that, matching the folk intuition, ellipsis provides

[4] In (**?**), this is formulated as a rule applying to epsilon terms, but, to equivalent effect, this process can be defined as an evaluation step on the paired restrictor specifications.

a window on the concept of context required in linguistic explanation. Context itself as we have seen, in DS, includes a record of parse states, primarily the most immediate one, and these are triples of word-sequence, (partial) tree structure which is the output of the processing of the words, and the set of actions that led to the build up of the structure. With this new concept of context (departing from regular denotational assumptions as in Dalrymple et al, Stalnaker etc), it is argued that the full range of semantic and syntactic effects displayed in ellipsis can be expressed while preserving a unitary base to the construal process.

For example, in (2), the strict interpretation can be established by presuming that the content of the predicate established for the first conjunct is reiterated, with the effect that the predicate is identically applied in both conjuncts, one form of *parallelism*:

(2) John checked over his mistakes, and so did Bill/Bill too.
'Bill checked Bill's mistakes' (sloppy)
'Bill checked John's mistakes' (strict)

The sloppy interpretation also can be established by identity with some construct taken from context, but in this case, the second parallellism effect, it is the sequence of actions taken that is identically applied in both conjuncts. The sequence of actions adopted in the first conjunct of (2) are (a) the actions introducing a predicate and (b) the actions introducing as object argument of that predicate a term whose restrictor is a relation between the entity under construction and some individual identified as the subject of the predicate. For the resolution of the ellipsis, it is this choice of actions that is selected from the first conjunct, and with this choice an interpretation exactly parallelling the mode of construal of the first conjunct can be built up which is nevertheless denotationally distinct from that of the first conjunct. For the hearer such selection yields an appropriate interpretation; for the speaker, it is because such selection yields the intended interpretation (his *goal tree*) that no more needs to be said. In being predicted to be available, *parallelism* effects in construal can thus be explained, while nevertheless capturing the diversity of interpretations apparently developed from a single antecedent source: what is reconstructed at the ellipsis site is reiteration of the very same actions used in building up interpretation for the first conjunct, applied now to the partial tree induced by the form of the fragment. Formally this involves having the ellipsis site be decorated with a metavariable of predicate type, $Ty(e \rightarrow t)$ constraining the choice of sequence of actions. In (2), there is only one such available sequence of actions, but in other cases, there may be more than one, leading to ambiguity:

(25) Sue was checking her results because she was worried her teacher was checking them and Molly was too.

In similar vein, syntactic constraints are also analysed in tree-growth terms. In particular, the Complex NP constraint debarring long-distance dependency across relative-clause boundaries is reconstructed as a constraint on updating the underspecified structural information projected by the relative pronoun:

(5) The man who Sue is worried that her sister is planning to marry lives in Austria.

(6) *The man who Sue is concerned about her sister who is planning to marry lives in Austria.

With this perspective on structural constraints, the parallel restriction on antecedent-contained ellipsis emerges unproblematically, for the linguistic content of the fragment itself provides all that is necessary to predict ellipsis construal sensitivity to such a constraint:

(3) John interviewed every student who Bill already had.

(4) *John interviewed every student who Bill ignored the teacher who already had.

The fragment in these cases is minimally made up of a determiner, a nominal and a relative pronoun initiating some clause which contains the ellipsis site. It is this sequence that determines the requisite type of construal. First, the noun content includes a variable which will form the restrictor of the quantificational term. Second, taking up the option of constructing a linked tree, a newly emergent tree is constructed in which an unfixed node is constructed, and the relative pronoun taken to decorate it with a second copy of this variable. It is then the weak domination relation associated with that unfixed node, ($\langle\uparrow_*\rangle ?Ty(t)$), which determines that its position must be resolved within the domain of a single tree. Whatever sequence of actions is selected as the construal of the ellipsis site has therefore to conform to this restriction, for this sequence of actions must extend the partial tree constructed from the fragment to yield a complete tree. In other frameworks, this would be a restriction articulated within the component of syntax, independent of any interpretation considerations. In DS, however, with syntax defined in terms of growth of representations of content, the restriction in question is expected to be imposed as a constraint on ellipsis construal, for the former is a constraint on tree growth which the latter, also a process of tree growth, is required to satisfy.

5 Fragments in Dialogue

Following through on this shift into a procedural perspective, **?** defined a concept of *wellformedness* with respect to context, opening the way for arbitrary fragments to be seen as wellformed as long as they occur in a particular environment. Under this definition, fragment construals and the context which they can extend can both be partial and dependent on the presence of each other for wellformedness. This provides a basis from which phenomena like (7)–(13) can be analysed using the same mechanisms for structure-building as made available in the core grammar. It has been noted that the range of interpretations these fragments receive in actual dialogue seems not to involve well-defined boundaries (see **?**). We suggest, nonetheless, that the grammar itself provides the mechanisms for processing and

integrating such fragments in the current structure, the dynamics of how information accrues in language processing being the core syntactic concept, even though the precise contribution of such fragments to the communicative interaction may need in addition pragmatic inferencing as a basis for selection between putative interpretations (**??**) CHECK appropriacy here/and 2009??). In DS, context includes a record of parse states, primarily the most immediate one, which both speaker and hearer can presume on. Hence for either speaker or hearer any reiteration of what is provided by the context may be unnecessary, given that the fragment provides sufficient trigger to make some appropriate selection from context recoverable. Modelling the transition between speakers as the transition between parse states means being able to capture the dynamics more directly via key aspects of the grammar. A distinct advantage is a continuum discerned from what are standardly seen as grammar-internal phenomena (eg ellipsis) to what are usually taken as distinct dialogue phenomena (fragments such as clarifications, extensions, reformulations and corrections). On this view, there simply is no essential difference: mechanisms for interpretation apply equally intra-sententially, inter-sententially, and across participants.

But further, the account unifies these in ways that shed light on the complex linking between contextual and linguistic information. Firstly, being able to directly draw on grammatical resources in this way means avoiding a model of interlocutor coordination via external mechanisms superimposed on a mode-neutral grammar formalism, since the key mechanisms involved are essentially internal to the grammar. Secondly, the account sheds new light on the grammar-parser contribution to disambiguation. Given this more fine-grained model of how interlocutors link current utterances with previous (discourse) contextual information, the familiar challenge of how to model multiple interpretive and structural options is opened up by the recurrent, often overlapping fragments, as in (12). Two features of DS are crucial in responding to this challenge: incrementality and parsing/generation bi-directionality. The incrementality of the DS approach, and the multiple parsing options made available by the DS framework might seem to increase complexity of the interpretive task in virtue of the direct expression of context-dependence. Nonetheless, the DS account enables one to express how interlocutors are able to exploit the inherent incrementality afforded by the grammar to manage such rapid increase in available options. Employing fragments incrementally in the build up of construal, hearers are able to immediately respond to a previous utterance at any point in the construction process, hence subsententially as well as sententially, so that interlocutors can constrain interpretation choices in an ongoing way – by clarification, acknowledgement, during the construction of even a single propositional formula. Such a possibility is not open to more conventional sentence-based frameworks where the locus of context dependency of linguistic processing is external to the core grammatical resources. Moreover, from a structural point of view, incremental resolution of parsing uncertainty through interaction reduces the complexity of multiple parsing paths opening up at each choice point: given that fragments are produced and processed at particular (sub-sentential) points allows for cutting down, at this particular point, the potential alternatives that have to be kept as available throughout the parse. Further, incrementality allows reduction of

uncertainty as regards the structural antecedent of the fragment since this fragment is processed while the pointer is located exactly at the intended antecedent node.

The parsing/generation bi-directionality of DS, in addition, makes it straightforward to model switching between speaker and hearer. Indeed in split exchanges, B's parse tree transparently reveals exactly where need of clarification or miscommunication may be arising, as it will be at that node from which a sub-routine extending it takes place. According to the DS model of generation, repeating or extending 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. In particular, 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). With this cycle of contribution-response-contribution, the effect of clarifications and the like, despite appearing to indicate misunderstanding, is in fact to narrow the focus to a specific point of query, enabling interlocutors the opportunity to make quite fine-grained adjustments to their own understandings.

6 Fragments in Dynamic Syntax

With these points in mind, we now turn to the DS account of fragment processing in dialogue. With context including the current partial tree, word sequence to date, and actions used to date, ellipsis construal can target any of these. In particular, given that both parsing and generation make use of the very same mechanisms of tree growth, split/joint utterance data are directly predicted. Switch from hearer to speaker is predicted to be possible at any arbitrary point in the dialogue without such fragments having to be interpreted as propositional in type (as is standard elsewhere, e.g. (**?**)) – the parser turned generator simply continues from the partial parse tree that has been established, relative to their own, possibly novel, ideas as to how that emergent tree should be completed; and the generator merely loses the initiative, but has a corresponding partial tree from which to process their respondent's attempt at completing it.[5]

6.1 Non-interruptive Fragments

We now have the basis for analysing extensions, non-repetitive acknowledgements, clarifications etc. which build on what has been previously said by way of confirming or requesting confirmation of the previous utterance. Recall examples (7) and (12), repeated below:

(7) A: Bob left.
B: the accounts guy, (yeah).

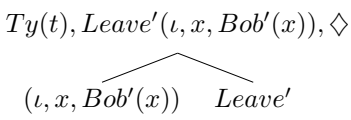
[5]Given the DS concept of *linked* trees projecting propositional content, we anticipate that this mechanism will be extendable to fragment construal involving inference (see e.g. **??**).

(12) 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 slide [unclear] on my heart. [BNC: KPY 1005-1008]

Fragments which reformulate an interlocutor A's utterance can be seen in two ways: either as (a) confirmations/extensions of A's utterance after the whole of her utterance has been integrated, see (7), or as (b) interruptions of her, A's, utterance, see (12), or . However, in DS, both are modelled in the same way: as incremental additions.

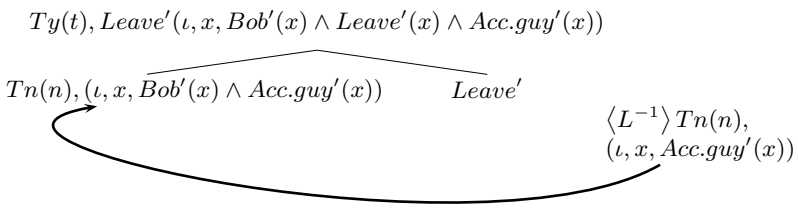
In (7), B's response *the accounts guy* constitutes both a reformulation of A's utterance, and an extension of A's referring expression, in effect providing the appositive expression 'Bob, the accounts guy'. B has presumably processed A's original utterance and achieved some identification of the individual associated with the name *Bob*: that is to say, B has constructed a full content representation for this utterance. In this case, as B's content corresponds with that intended by A, and the resulting *linked* trees are therefore consistent with each other, the reformulation has the effect of acknowledgement. In DS terms, B's context after processing A's utterance contains the following tree:

(26) B's Context:[6]



B can now re-use this contextual representation as the point of departure for generating the expression *the accounts guy*, using the same apposition mechanism as defined in Section **??**. 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:

(27) B's goal tree for uttering *the accounts guy*:

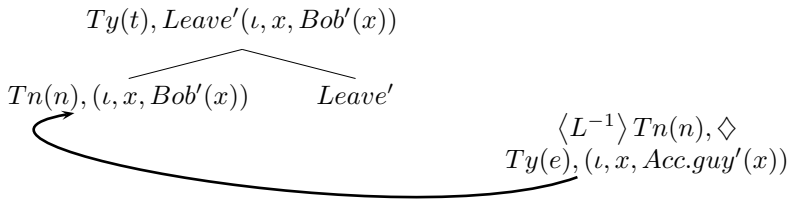


In order for B to produce the fragment relative to the content of what A has said (the immediate context of his own utterance), he has to test-parse the fragment first and incrementally check that the subsumption relation is preserved between each

[6]Items like *yeah* have a metacommunicative function in dialogue (*backchannels*) and are not currently included as part of the main DS propositional content.

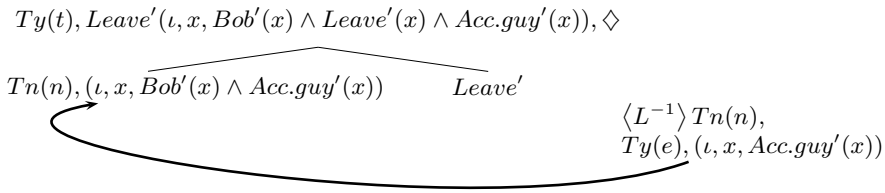
parsing transition and his goal tree. Simplifying for illustration purposes, the (test-) parsing steps include attaching a *linked* tree to the correct node of that context tree, moving the pointer there and then processing the content of the apposition. If all these steps subsume the goal tree license to produce the words will ensue. So, in B's responding utterance, firstly a *link* transition is introduced, and the pointer is moved to the newly constructed tree to which the iota term provided by *the accounts guy* can be added as decoration as a test-processing step, from which B can verify that the resulting tree will subsume his goal tree in (27):[7]

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



Updating this representation according to the DS processing protocol involves adding the acquired restrictions at the node from which the *linked* tree is projected to create the composite term, finally passing the information up to the top node of the main tree. This completes the test-parse and, given that matching with the goal tree is verified, the NP *the accounts guy* can be uttered:[8]

(29) B's test-parse tree licensing utterance of *the accounts guy*:



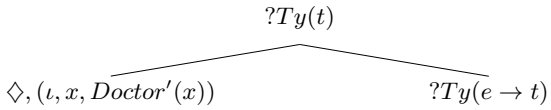
6.2 Interruptive Clarification

In the acknowledgement case (7), the tree relative to which the *linked* structure is built is *complete*; but the very same mechanism can be used when the interlocutor needs clarification, and the tree being built is still *partial*. In (12), B has built only a partial tree at the point of interruption:

(30) B's CONTEXT:

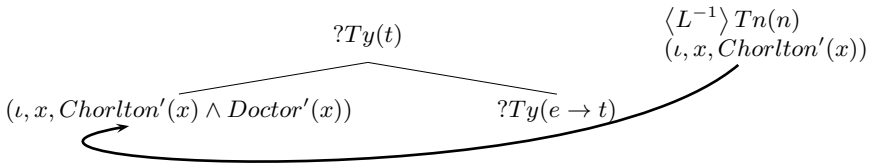
[7] Given that for reasons of space we do not show the internal structure of epsilon/iota terms on the graphics the presentation is simplified.

[8] In fact, given the incrementality of DS, each single word is uttered individually upon the subsumption check but we suppress these steps here for simplicity.



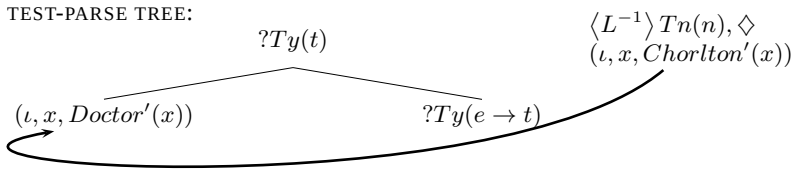
In order to request clarification of the intended referent of *the doctor*, B again takes as his goal tree a tree decorated with an expansion of the term constructed from parsing A's utterance, as a means of supplying more specific information to aid in the task of identifying who is being talked about. The fact that this time B's goal tree is partial (he has not completed the parse of a full proposition before asking for clarification) causes no problem for the analysis:[9]

(31) B's GOAL TREE:



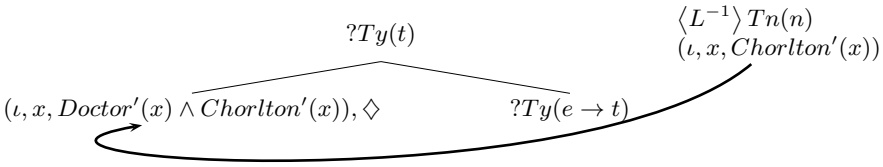
Using the very same mechanism as in (7) of building a *linked* structure constrained to induce shared terms, B can generate (and A can parse) the name *Chorlton*, no matter that neither has completed the parse tree for A's original (unfinished) utterance. This name, contributing a term with the restrictor that the individual picked out must be named 'Chorlton', is used to decorate the *linked* node. As this subsumes the goal tree of (31), the name *Chorlton* is licensed to be uttered:

(32) B's TEST-PARSE TREE:



The outcome of this process, when the *linked* structure is evaluated, is a composite term $(\iota, x, Doctor'(x) \wedge Chorlton'(x))$ at the node at which the *linked* tree was attached, extending the initial iota term:

(33) B's TEST-PARSE TREE:



This process, therefore, is identical to that employed in B's utterance in (7), 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 derived from B's utterance has led to (see (**?**) for discussion). 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.

[9] We ignore here any representation of question-hood, since our emphasis is on common mechanisms. See **?** for preliminary discussion.

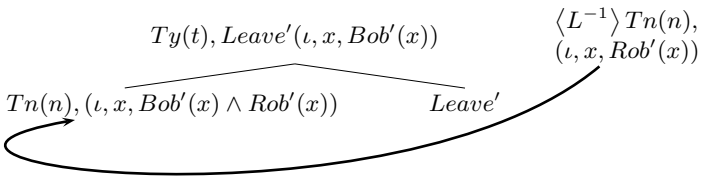
6.3 Correction

It might be argued nonetheless that correction is intrinsically a dialogue phenomenon; and indeed the recognition that some information is inconsistent with what one has processed constitutes an inferential step requiring access to the parser's more general knowledge base. Suppose that B mishears and requests confirmation of what he has perceived A as saying, but that he is mistaken; and A in turn rejects B's utterance and provides more information:

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

Of course, A can process B's clarification request in exactly the same way as set out in Section 6.2 above, as an extension of her own context via *linked* tree construction. This leads to a representation as follows:

(34) A's PARSE TREE:



In order for A to establish that the information here leads to inconsistency (the set denoted by $Bob'(x) \wedge Rob'(x)$ is empty), the individual so described not having two such names), she has to be able to retrieve information independently assumed by her (again this too needs a specified interface with an inference model that is not provided here). But assuming that this is available, the tree can be recognised as specifying information that is inconsistent, which would lead to rejection. Rejection is therefore analysed here as simple disagreement: B's utterance has been understood, but simply judged as incorrect.

To generate her subsequent correction, A need only establish as the current most recent representation in context her original goal tree (the most recent *consistent* tree available). This can be monotonically achieved by recovering and copying this original goal tree to serve as the current most immediate context. Generating the corrective *Bob, the accounts guy* then proceeds exactly as in the previous sections. Note that this is not a case of B's utterance content being (non-monotonically) *removed* from the context, even for A: corrected representations must be maintained in the context as they can provide antecedents for subsequent anaphoric expressions, as in:[10]

(35) A: Bob left.
B: Rob?
A: No. HE'S in Beijing these days. Bob, the accounts guy.

[10] It is notably harder to recover ellipsis construal appropriately across an intervening utterance, but it is by no means impossible (see **?**).

7 Summary and Evaluation

As these fragments and their construal have demonstrated, despite serving distinct functions in dialogue – as acknowledgements, corrections, extensions, etc – and despite the fact that additional inferences might be needed to confirm threatening inconsistency, the mechanisms which give rise to the distinct functions are nonetheless general strategies for tree growth that are independently available. In closing, it is of some interest to reflect that, even though DS is a grammar formalism and so not in principle providing a full theory of either utterance understanding or dialogue interactivity, nevertheless, there are clear predictions which the account would lead one to expect. 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*: hence its prevalent use in conversational dialogue is expected. It is not merely that fewer words are used in such minimal utterances, hence preferred on a trivial cost basis, but that in such elliptical re-iteration, a whole sequence of, in principle, independent production/parsing choices is taken to be fixed by the way in which such decisions were taken in the processing of the antecedent string. Such a determinism means that, with information culled from context, representations do not have to be constructed afresh via costly processes of lexical retrieval, choice of alternative parsing strategies, etc.

A further quandary in dialogue construal is that, no matter what avenues for economising their efforts interlocutors may make use of, the hearer is nevertheless faced with an increasing set of interpretative options at any point during the construction of representations. The hearer may choose to delay a disambiguating move until further input potentially resolves the uncertainty, putting the uncertain node unresolved in memory. However, as further input may not help resolve this uncertainty but only increase the available options cumulatively —a point may arise after another on which the hearer needs clarification—, maintenance of these open options becomes difficult for a human processor. Again, despite the DS objective of providing merely a formalism that licenses possible tree-growth transitions, 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 (12) 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 (13)). According to the account offered here, the dynamics of interaction allow fully incremental generation and integration of fragmental utterances so that interlocutors can constantly provide direct evidence of each other's representations, thereby being better able to immediately make necessary micro-adjustments. So incorporating a reflection of time-linear dynamics in the grammar formalism itself not merely narrows the competence-performance gap by definition, but it serves in part to directly address the complexity issues normally taken to be a performance consideration, an evaluation metric on parser/generator design. [I

THINK SOMETHING LIKE THIS IS NEEDED OTHERWISE THE CONCLUSION ISN'T MATCHING THE GOAL OF THE PAPER AS INITIALLY SET OUT]

Finally, in modelling the transition between speakers as the transition between parse states defined grammar-internally, fundamental aspects of dialogue are modelled in a largely mechanistic manner (following **?**: (**??**, see also)), a move which is echoed in emerging results in cognitive science more generally.[11] This opens up the possibility of characterising language as a set of mechanisms for communicative interaction without any need of recourse to high-level representation or expectation of other people's beliefs.

Acknowledgements

This work was supported by grants ESRC RES-062-23-0962 and Leverhulme F07 04OU, and reflects ongoing work. We are grateful for ongoing feedback to Patrick Healey, Gregory Mills, Arash Eshghi, Wilfried Meyer-Viol, Ronnie Cann, Graham White. Mistakes however have to be seen as our own.

[11] There is a range of results linking action and perception within a common framework (e.g. **??**), on how various cognitive mechanisms for "sharing" representations may facilitate joint action (**?**), on research into common representations underlying both speaking and hearing (e.g. **??**), and on imitation as *behavior parsing* **??**. Neuroscientific research on such parity between action and perception is now well-established (eg Neural Networks 2006 'Special Issue on The Brain Mechanisms of Imitation Learning'), as is work on the role of such mechanisms in communication (e.g. **?**).