Conversational Interactions: Capturing Dialogue Dynamics

Arash Eshghi, Julian Hough, Matthew Purver (QMUL, London)
Ruth Kempson, Eleni Gregoromichelaki (KCL London)

1 Background

A common position in the philosophy of language has been the separation of the 'intentionality' of natural language (NL) and thought from the exercise of the capacities and epistemic resources that underpin perception and action. From this point of view, an adequate theory of meaning is given in a formal theory of 'truth' for NL (see e.g. Davidson (1967); Larson and Segal (1995); Montague (1970)). Such a theory for NL provides a systematic account of the finite system of resources that enables the user of the theory to understand every sentence of the language. However, when we turn to examine the employment of this knowledge in realistic settings, i.e. in communication, it was believed that by stepping outside this methodology, we would inevitably be led to have "abandoned not only the ordinary notion of a language, but we have erased the boundary between knowing a language and knowing our way around in the world generally." (Davidson, 1986). As a response to this danger, until recently, a common methodology in Theoretical Linguistics has been "to try to isolate coherent systems that are amenable to naturalistic inquiry and that interact to yield some aspects of the full complexity. If we follow this course, we are led to the conjecture that there is *a generative procedure* that "grinds out" linguistic expressions with their interface properties, and *performance systems* that access these instructions and are used for interpreting and expressing one's thoughts" (Chomsky, 2000, 29, emphasis ours). This methodological principle that dictates strict separation of the (modelling of) linguistic knowledge (competence) and the application of this knowledge in actual situations of language use (performance) has been called into question recently by several researchers interested in modelling the capacities underpinning NL use. Contrary to the standard "autonomy of syntax" hypothesis, grammatical models have recently begun to appear that reflect aspects of performance to varying degrees (Hawkins, 2004; Phillips, 2003; Lombardo and Sturt, 2002; Ginzburg and Cooper, 2004; Kempson et al., 2001; Cann et al., 2005b; Ginzburg, 2012). One type of motivation for this shift is that a number of researchers have recently pointed out that a range of metacommunicative acts (in track 2: Clark (1996)) running in parallel with the communicative acts (in track 1) have to be characterised as part of the grammar itself (e.g. Purver (2006); Fernández (2006); Ginzburg (2012); Gregoromichelaki et al. (forthcoming)). Another type of motivation, espoused by *Dynamic Syntax* (DS, Kempson et al. (2001); Cann et al. (2005b)), is the demonstration that standard syntactic phenomena can be explained in a cognitively non-arbitrary fashion by taking a fundamental feature of real-time processing - the concept of underspecification and incremental goal-directed update - as the basis for the formulation of syntactic constraints.

In the domain of semantics and pragmatics, there has long been work emphasising the role of underspecification in the derivation of meaning and formulating notions of 'procedural meaning' that cannot be accommodated under the truth-theoretic conceptions of semantics (see e.g. Sperber and Wilson (1995); Levinson (2002)). Further inadequacies of traditional semantic theories have been further highlighted by the pioneering work of Robin Cooper and colleagues who, along with DRT and related frameworks, have drawn attention to the importance of formalising the contribution of an extended, structured notion of the (multi-modal) context in supplying an adequate theory of interpretation for NLs. In this attempt to provide an adequate theory of language understanding, attention has shifted away from a strict formulation of a truth theory to the modelling of the structure of the information manipulated during perception and action as it interfaces with linguistic processing (see e.g. Larsson (2011)). Inspired by work in Situation Semantics and DRT the most recent formulation of this effort has been via the employment of Type Theory with Records (TTR), a transparent representation format allowing the specification and seamless interaction of multiple types of information. In recent years this has led to a significant expansion of the data deemed appropriate for inclusion in a formal theory of interpretation, namely, the modelling of the use of language in interaction and the demands that this places on appropriate semantic models (see e.g. Ginzburg and Cooper (2004); Ginzburg (2012)).

In this paper, we set out the case for combining Dynamic Syntax (DS, Kempson et al., 2001; Cann et al., 2005b) and the Type Theory with Records framework (TTR, Cooper, 2005) in a single model (DS-TTR) in order to capture what is in our view the most fundamental aspect of linguistic knowledge, namely, its exercise in interactive settings like conversational dialogue. DS is an action-based formalism that specifies the 'know-how' that is employed in linguistic processing, in contrast to standard formalisms which codify (specifically linguistic) propositional knowledge of rules. At the heart of the DS approach is the assumption that grammatical constraints are all defined in terms of the progressive growth of representations of content, with partial interpretations built step-by-step during interaction with context on a more or less word-by-word basis. In consequence, DS is well-placed to provide the basis for a fine-grained integrational model of language use that incorporates various aspects of the interface with perception and action in a single representation. The data we present below show that such a model is required to account for the syntactic properties of various phenomena that arise as a result of language use in interaction, instead of characterising them as "dysfluencies" or "performance phenomena", hence outside the remit of core grammar or truth-theoretic characterisations. The representations required for the modelling of such phenomena can be provided in a straightforward manner by the TTR framework which allows for the fine-grained incrementality appropriate for showing how such representations can be progressively established and which is at the heart of what DS is committed to capturing. The basis for this is the recursive nature of the TTR records and record type format through its notion of subtyping. This allows the specification of underspecified objects, through partially specified types, which can be progressively specified/instantiated as more information becomes available. As a result, the formulation of highly structured models of context, where uniform representations of multiple types of information can be supplied and their interaction modelled, becomes achievable (see e.g. Larsson, 2011). In addition, TTR employs a general type-theoretic apparatus with functions and function types so that standard compositional lambda calculus techniques are available for defining interpretations, thus capturing the systematicity and productivity of linguistic semantic knowledge. When combined with a grammar formalism in which "syntax" itself is defined as driving incremental growth of interpretation, strict word-by-word incrementality of semantic content representations becomes definable, enabling the maximum amount of semantic information to be extracted from any partial utterance and represented as a record type to which fields are added incrementally as more words are processed in turn. Furthermore, inference, as one of a range of operations, is definable over these sub-propositional record types, so that TTR is particularly well suited for representing how partial semantic information is step-wise accumulated and exploited. And because types can be treated as objects in their own right, it also becomes possible to integrate the reification and manipulation of both contents and grammatical resources for metarepresentational/metalinguistic purposes.

In sum, as we will demonstrate in what follows, the combination of these two components, DS and TTR, opens up the means of characterising phenomena that go far beyond the data expressible within standard syntactic and semantic theories. We also show that such phenomena cannot be handled without radically modifying the competence-performance distinction as standardly drawn. The more orthodox view, as we will show, far from being a harmless abstraction that will eventually seamlessly integrate with a unified explanation of the capacities that underpin language use, turns out to have provided a distorted view of the NL phenomenon, resulting in a misleading formulation of the nature of knowledge required for understanding and production in realistic settings (for philosophical arguments supporting this view see also Millikan (2004); McDowell (1998)).

2 The scope of grammar

2.1 Linguistic knowledge: the standard view and the view from the DS-TTR perspective

Standardly, the formulation of grammars abstracts away from use as it is assumed that use of language is an operation that must have at its core propositional knowledge of an independently specifiable syntactic theory and a theory of meaning. Syntax is confined to the licensing of sentence-strings and so delimiting the set of well-formed sentences of the language; and semantics is then defined as the application to that set of structured strings of a truth theory yielding propositions as denotations, this being the interface point at which the contribution of grammar stops and pragmatics takes over. As a consequence of this stance, classical truth-based semantic theories have enshrined Frege's *Context Principle* which holds that one should "never ask for the meaning of a word in isolation, but only in the context of a proposition" (see e.g. Davidson (1967)). Under such a view, it is only as they play a role in whole sentences that individual words or phrases can be viewed as meaningful. One of the reasons behind this stance is that the basic units of linguistic understanding are taken to be *propositional speech acts*, as the minimal moves in conversation, and steps of inference as expressed via either classical logical calculi or inductive generalisations are invariably modelled as involving propositions as premises and conclusions. Since most standard pragmatic models take such inferences as the basis for explaining communication, via complex propositional reasoning regarding propositional attitudes like speaker intentions, it is a requirement that the grammar delivers such objects as input to further pragmatic processing. Given this standard view of grammar, as independent of language use, even (psycholinguistic) models within the language-as-action tradition bifurcate the concept of 'language' as language$_u$ (that is, language-in-use), to be distinguished from language$_s$ (that is, language structure)- see e.g. Clark (1996).

In contrast, the procedural architecture of Dynamic Syntax (DS) that models "syntax" as "knowledge-how" incorporates into the grammar two features usually associated with parsers, namely, incrementality and fine-grained context-dependence. These features are argued to constitute the explanatory basis for many idiosyncrasies of NLs standardly taken to pose syntactic/morphosyntactic/semantic puzzles (see Cann et al. (2005a); Kempson and Kiaer (2010), and papers in Kempson et al. (2011b), also Chatzikyriakidis and Kempson (2011)). This revision of what kind of knowledge a 'grammar' encapsulates is appropriate for combining it with some of the foundational assumptions that underlie the employment of TTR by Cooper and colleagues, namely, the provision of an integrated architecture that handles the integration of perception and action in language use. As standard in TTR, the semantic contribution of utterances can be taken as operations on fine-grained structured representations of contexts but, extending its expressivity, the incorporation of incrementality within the grammar formalism is justified by the application of the grammar-defined principles to a much broader remit of data than is possible in conventional grammars, in particular to include the rich set of data displayed in conversational exchanges.

2.2 Incrementality, radical context-dependence and dialogue phenomena

2.2.1 The (non-)autonomy of syntax

In conversation, evidence for incrementality is provided by the fact that, as can be seen in (1), dialogue utterances are fragmentary and subsentential, yet, intelligible actions can be performed all the same in the context of the ongoing interaction with interlocutors and the physical environment:

(1) Context: Friends of the Earth club meeting
A: So what is that? Is that er... booklet or something?
B: It's a [[book]]
C: [[Book]] (*Answer/Acknowledgement/Completion*)
B: Just ... [[talking about al you know alternative]] (*Continuation*)
D: [[On erm... renewable yeah]] (*Extension*)
B: energy really I think... (*Completion*)
A: Yeah (*Acknowledgment*) [BNC:D97]

Moreover, this interactivity is buttressed by the ability of the participants to manifest their progressive understanding as they "ground" each other's (subsentential) contributions through *back-channel* contributions such as *yeah*, *mhm*, etc (see e.g. Allen et al. (2001)). Moreover, the placing of items like inserts, repairs, hesitation markers etc. far from being "errors (random or characteristic) in applying knowledge of language in actual performance" (Chomsky, 1965, p. 3), follows systematic patterns that show subtle interaction with grammatical principles at a sub-sentential level (Levelt, 1983; Clark and Fox Tree, 2002):

(2) well, . I mean this . uh Mallet said Mallet was uh said something about uh you know he felt it would be a good thing if u:h . if Oscar went, (1.2.370)

This implies that dialogue phenomena like self-repair, interruptions, extensions, corrections etc. require modelling of the participants' incremental understanding/production; and if, as we will show, particular NL grammars are required to provide the licensing of such constructions then such grammars need to be equipped to deal with partial/non-fully-sentential constructs. Modular approaches to the grammar/pragmatics interface deny that this is an appropriate strategy. Instead they propose that the grammar delivers underspecified propositional representations as input to pragmatic processes that achieve full interpretations and discourse integration (see e.g. Schlangen (2003), following an SDRT model). However, an essential feature of language use in dialogue is the observation that on-going interaction and feedback shapes utterances and their contents (see e.g. Goodwin (1981), Clark (1996), among many others), hence it is essential that the grammar does not have to licence whole propositional units for semantic and pragmatic evaluation to take place. And this is the strategy DS adopts as it operates with partial constructs that are fully licensed and integrated in the semantic representation immediately. This has the advantage that online syntactic processing can be taken to be implicated in the licensing of fragmentary utterances, even when these are spread across interlocutors (*split utterances*), without having to consider such "fragments" as elliptical sentences (Merchant, 2004), or as contributing pragmatically derived propositional contents (Stainton, 2006) or non well-formed in any respect. And this is essential for a realistic account of dialogue as people can seamlessly take over from each other in conversation. They may seek to finish what someone else has in mind to say as in (3), but equally, they may interrupt to alter what someone else has proffered, taking the conversation in a different or even contrary direction, as in (4) and (5) :

(3) Gardener: I shall need the mattock.
Home-owner: The...
Gardener: mattock. For breaking up clods of earth.[BNC]

(4) (A mother, B son)
A: This afternoon first you'll do your homework, then wash the dishes and then
B: you'll give me £10?

(5) (A and B arguing:)
A: In fact what this shows is
B: that you are an idiot

Furthermore, this is a form of exchange that children can join in from a very early age to complete someone else's utterance, as witness the English nursery game Old MacDonald had a Farm:

(6) A: Old MacDonald had a farm. E-I-E-I-O. And on that farm he had a
B: cow.
A: And the cow goes
B: Moo.

And carers may trade on this ability, e.g. in talking in a nursery class, again from a very early age:

(7) (teacher to first of a set of children sitting in a circle)
A: Your name is ...
B: Mary
A: (turning to next child) And your name is...
C: Susie
and so on

A respondent, as in (6) and (7) the child, may just complete a frame set out by the dialogue initiator. However, commonly, participants may, in some sense, "just keep going" from where their interlocutor had got to, contributing the next little bit:

(8) A: We're going to London
B: to see Granny?
A: if we have time.

Such exchanges can indeed be indefinitely extended, so that each contributor may only be contributing some intermediate add-on without either of them knowing in advance the end-point of the exchange:

(9) (a) A: Robin's arriving today
(b) B: from?
(c) A: Sweden
(d) B: with Elisabet?
(e) A: and a dog, a puppy and very bouncy
(f) B: but Robin's allergic
(g) A: to dogs? but it's a Dalmatian.
(h) B: and so?
(j) A: it won't be a problem. No hairs.

The upshot of this is that it is often hard to tell where one sentence begins and the next starts. Does the exchange in (9) consist just of one sentence or perhaps two - i.e. all that precedes "problem" plus the final "No hairs"? Or does it consist of one sentence for each fragment as individually uttered?

Faced with this kind of dilemma, it might be tempting to dismiss the phenomenon altogether as a dysfluency of conversational dialogue, but the problem is not merely one of incompleteness in characterisation of a single sub-area of language use. The form of these "fragments" is not random but, to the contrary, follows the licensing conditions specified by the NL grammar – syntactic dependencies of the most fundamental sort hold between the subsentential parts:

(10) A: I'm afraid I burned the buns.
B: Did **you** burn
A: **myself**? No, fortunately not.

(11) A: D'you know whether **every waitress** handed in
B: **her** taxforms? A: or even **any** payslips?

Given that the standard motivation behind the sententialist/propositionalist bias in syntax and semantics is the assumption that only sentences/propositions can be used in the performance of speech acts, it might seem but a minor extension to include such phenomena under some propositional/sentential reconstruction with additional encoded or inferred speech act specifications. This strategy has been applied in many cases of ellipsis where either an underlying sentence is constructed and in greater part deleted (e.g. Merchant (2004)) or in recent models of dialogue where a speech act specification and a propositional content are constructed by operations on context (see e.g. Ginzburg (2012)). However, the phenomenon is much more general than such analyses suggest. People can take over from one another at any arbitrary point in an exchange, setting up the anticipation of possible dependencies to be fulfilled. We have already seen that it can be between a preposition and its head, (9b-c), between a head and its complement (9f-g), between one conjunct and the next (9d-j) etc. (10) involves a split between a reflexive pronoun and its presented antecedent. (11) involves a split between a quantifying expression and some pronoun that it binds, and then across a disjunction and another shift of speakers to a negative polarity item dependent on that initially presented quantifier. (3) involves a split between determiner and noun. The upshot is that switch of participant roles is possible across ALL syntactic dependencies (Purver et al., 2009): participants in a dialogue seem, in some sense, to be able to speak with a single voice, even while yet directing the conversation as they individually wish. Unless the grammar reflects the possibility of such dependencies to be set and fulfilled across participants (or, in fact as we will see below (12)-(17) in interaction with the physical environment) not a single grammatical phenomenon will have successfully been provided with a complete, uniform characterisation.

On the other hand, any attempt to reflect this type of context-dependence, and the attendant sense of continuity it gives rise to, through grammar-internal specifications will have to involve constraints on fragment construal that go well beyond what is made available in terms of denotational content: indeed such constraints will have to include the full range of syntactic and morphosyntactic dependencies. As Ginzburg and Cooper (2004); Ginzburg (2012) observe (following similar observations in Morgan (1973, 1975)), in all case-sensitive languages there is sensitivity of "fragment" expressions to some notion of recoverable antecedent syntactic environment, so that invariably the fragment uttered has to match the morphosyntactic requirements set by the expression to which it is providing an extension. For all such cases, a semantic/pragmatic characterisation on its own will not be sufficient, and syntactic licensing is essential. However the puzzle is not yet complete. As has already been demonstrated by Stainton (2006), speakers can perform genuine speech acts via use of subsentential constituents without needing first to recover complete syntactic sentences or sentence contents. Nevertheless, contrary to Stainton's assumptions, such "fragments" too need to respect the morphosyntactic requirements of the relevant NL, a fact indicating the employment of the grammar at a subsentential level, even when derivation of the speech act content is achieved purely pragmatically:

(12) Context: A and B enter a room and see a woman lying on the floor:
A to B: Schnell, den Arzt/*der Arzt [German]
"Quick, the doctor$_{ACC}$ /*the doctor$_{NOM}$"

But these data are also problematic for any account that analyses such phenomena in purely linguistic terms by defining rules that make reference to (covert) antecedent *utterances* with some specified NL syntactic form (see e.g. Ginzburg (2012)). For, like anaphora and standard elliptical phenomena, most dialogue phenomena, clarifications, extensions, corrections etc. can occur without linguistic antecedents:

(13) A is contemplating the space under the mirror while re-arranging the furniture and B brings her a chair:

tin karekla tis mamas?/*i karekla tis mamas? Ise treli? [Greek] [clarification]
the chair of mum's$_{ACC}$/*the chair$_{NOM}$ of mum's. Are you crazy?

(14) A sees Bill entering the building and turing to C exclaims:
A: o Giannis(?)/*ton Gianni(?) [Greek] [assertion(/clarification)]
the John$_{NOM}$/*the John$_{ACC}$

(15) A is looking for her keys and B points to a desk:
B: your desk? I've looked there. [clarification]

(16) A is handing a brush to B:
A: for painting the wall? [clarification]

(17) A is pointing to Bill:
B: No, his sister [correction]

Hence accounts that rely on rules that require reference to some salient linguistic form of utterance antecedent are not general enough for the phenomena at hand. What is needed, in our view, are representations of both (linguistic) content and context in which multiple (multi-modal) sources of information are all expressed in a single format. This will enable the modelling of linguistic resources that can make reference to and modify such representations in an incremental manner. On the DS-TTR account, as we shall see, morphosyntactic particularities, for example, do not warrant distinct levels of explanation in the update mechanisms needed for fragment construal, for morphological information, like all other aspects of morphosyntactic and syntactic specification is defined in terms of the constraints the morphological form imposes on appropriate integration into a structured content representation, effecting a specified update. Seen from this perspective, these dialogue data, far from being set aside as beyond the reach of grammar, in fact demonstrate how the grammar needs to be equipped with fine-grained licensing mechanisms that operate at the subsentential level with sensitivity to the time linear progress of interaction between the agents and the evolving context in which their interaction takes place.

2.2.2 Pragmatic/semantic "competence" and radical context-dependence in dialogue

These data are significant for pragmatics also. There has been an assumption held by almost all those working in pragmatics that the supposedly isolatable sentence meaning made available by the grammar should feed into a theory of performance that explains how, relative to context, such "sentences" can be uttered on the presumption that the audience will come to understand the propositional content which the speaker has (or could have) in mind. However, participants may well understand what each other is saying and switch roles well before any such propositional content could be interpreted to constitute the object relative to which some agent or other could hold a propositional attitude. These switches take place, recall, at any arbitrary point in the constructive process. Intentions of the parties to the dialogue may only emerge/develop during the exchange (Mills and Gregoromichelaki, 2010), and so cannot be intrinsic to all processes of communicative understanding, as is so generally assumed, for example, in the only existing formal model of completions, that of Poesio and Rieser (2010):

(18) A: Oh. They don't mean us to be friends, you see. So if we want to be . . .
B: which we do
A: then we must keep it a secret. [natural data]

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

(20) M: It's generated with a handle and
J: Wound round? [BNC]
M: Yes, wind them round and this should, should generate a charge

There is negotiation here, as to what is the best way to continue a partial structure as proffered by either party, with intentions of either party with respect to that content, possibly only emerging as a result of the negotiation. Utterances may also be multi-functional, so that more than one speech act can be expressed in one and the same utterance:

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

(22) Lawyer: Do you wish your wife to witness your signature, one of your children, or..?
Customer: Joe.

So there is no single proposition or indeed speech act that the individual speaker/hearer may have carried out.

The commitment to the recovery of propositions or propositional speech act contents as a precondition for either successful linguistic processing or effective interaction (Grosz and Sidner (1986)) has therefore to be modified; and so too does the presumption of there having to be explicit plans/intentions on the part of the speaker (Poesio and Rieser, 2010; Carberry, 1990). To the contrary, these data provide evidence that the grammar itself and its mechanisms can be exploited by all participants in a conversation as the means to progress that interaction at a subsentential level before any such speech-act or propositional content becomes available. In fact, in many cases, the participants can simply rely on the setting up of grammatical dependencies and the parallel in both speaker and hearer inducement to fulfil them in order to perform possibly composite speech acts (*grammar-induced speech acts*, Gregoromichelaki et al. (forthcoming)) without even requiring steps of inference or recovery of propositions (see also (7) and (11) above):

(23) Jim: The Holy Spirit is one who ...gives us?
Unknown: Strength.
Jim: Strength. Yes, indeed. .... The Holy Spirit is one who gives us? .....
Unknown: Comfort. [BNC HDD: 277-282]

(24) George: Cos they [unclear] they used to come in here for water and bunkers you see.
Anon 1: Water and?
George: Bunkers, coal, they all coal furnace you see, ... [BNC, H5H:59-61]

(25) A: And you're leaving at ...
B: 3.00 o'clock

(26) Therapist: What kind of work do you do?
Mother: on food service
Therapist: At ...
Mother: uh post office cafeteria downtown main point office on Redwood
Therapist: Okay [Jones & Beach 1995]

Such cases show, in our view, that "fragmentary" interaction in dialogue should be modelled as such, i.e. with the grammar defined to provide mechanisms that allow the participants to incrementally update the conversational record without necessarily having to derive or metarepresent propositional speech act contents or contents of the propositional attitudes of the other participants as the *sine qua non* of successful communication. In the exercise of their grammatical knowledge in interaction, participants justify Wittgenstein's view that "understanding is knowing how to go on". Metacommunicative interaction is achieved implicitly in such cases via the grammatical mechanisms themselves without prior explicit commitment to deterministic speech-act goals, even though participants can reflect and reify such interactions in explicit propositional terms (see e.g. Purver et al. (2010)), should they so choose. The fact that such reifications are possible, even though it requires that the dialogue model should provide the resources for handling them when they are explicit, does not imply that they operate in the background when participants engage in (unconscious, sub-personal) practices that can be described from the outside in explicit propositional terms. The level of explanation for explicit descriptions of actions and implicit practices is not the same. In parallel with Brandom's (1994) conception of the logical vocabulary as the means which allows speakers to describe the inferential practices that underlie their language use, conversational participants manifest their ability to "make explicit" the practices afforded to them implicitly by subpersonal procedures when either communication breaks down or when they need to verbalise/conceptualise the significance of their actions (for a similar view account of practices at other higher levels of coordination see Piwek (2011)).

The problem standard syntactic theories have in dealing with dialogue data can be traced to the assumption that it is sentential strings that constitute the output of the grammar, along with the attendant methodological principle debarring any attribute of performance within the grammar-internal characterisation to be provided. The semantic literature, on the other hand, focuses on the assumption that NL meaning can be modelled through a Tarski-inspired truth theory for NL. Neo-Davidsonians (e.g. Larson and Segal, 1995) further assume that knowledge of language consists in tacit propositional knowledge of the truth theory; this tacit knowledge is what enables individuals to produce and interpret speech appropriately in interaction with others with the same tacit knowledge. However, Davidson himself acknowledges that the individualistic perspective on what this knowledge consists in is inadequate:

> ... there must be an interacting group for meaning –even propositional thought, I would say– to emerge. Interaction of the needed sort demands that each individual perceives others as reacting to the shared environment much as he does; only then can teaching take place and appropriate expectations be aroused. (Davidson, 1994)

In this respect, Cooper and colleagues (see e.g. Ginzburg (2012) who employs Cooper-inspired TTR methods) have achieved the significant advance of defining an explicit semantic model that does not restrict itself to the modelling of informational discourse but instead attempts to describe the fine-grained structure of conversational exchanges that result in participant coordination (see e.g. Pickering and Garrod (2004)) and explores the ontologies required in order to define how speech events can cause changes in the mental states of dialogue participants. But, following standard assumptions, syntax is defined independently and in effect statically (however, see Ginzburg (2012, ch. 7)) which, in our view, prevents the modelling of the fine-grained incrementality observable in the split-utterance and repair data, a lacuna which the DS-TTR combination aims to repair. As we will see below, when embedded in the action-based incremental architecture provided by DS, the view of the semantic landscape changes. The instrumentalist Davidsonian stance towards the content assigned to subsentential constituents, as subordinate to sentential contents, has to be revised in that subsentential contributions provide the locus for as much and as significant (externalised) "inference" and coordination among participants as any propositional contributions. And to explain the relation between a provided partial structure as context and an update that completes or extends it, the concept of context has to be structural to a level of granularity matching that of syntax.

3 DS-TTR for dialogue modelling

In turning to the modelling of conversational dialogue, we will need concepts of incrementality applicable to both parsing and generation. Milward (1991) sets out two key concepts of *strong incremental interpretation* and *incremental representation*. These concepts apply to semantic incrementality, largely. Strong incremental interpretation is the ability to make available the maximal amount of information possible from an unfinished utterance as it is being processed, particularly the semantic dependencies of the informational content (e.g. a representation such as $\lambda\, x.like'(john', x)$ should be available after processing "John likes"). Incremental representation, on the other hand, is defined as a representation being available for each substring of an utterance, but not necessarily including the dependencies between these substrings (e.g. having a representation such as $john'$ attributed to "John" and $\lambda\, y.\ \lambda\, x.like'(y, x)$ attributed to "likes" after processing "John likes"). But there are three further concepts pertaining to incrementality to bear in mind. In order to express the incrementality intrinsic to syntax, we need to stipulate that these abilities exhibit *word by word incrementality*, whereby all information affiliated with a word must be taken to update the structure to which it applies as input immediately, whether structural, conceptual, or, if applicable, semantic. As we shall see, it is this third notion which lies at the core of Dynamic Syntax, in which syntax is defined in terms of such structural update. Furthermore, in order to model compound contributions as described in the examples given thus far, it also becomes evident that the representations produced by parsing and generation should be *interchangable*, as will be discussed in section 4.1. Finally, the notion of an incrementally constructed *procedural context* becomes important for modelling self-repair, a quality of the DS framework described in section 3.2 and exploited in section 4.2.

3.1 Combining Dynamic Syntax and TTR

Dynamic Syntax is a parsing-directed grammar formalism,

Dynamic Syntax (DS, Kempson et al., 2001) is a grammar framework which models the word-by-word incremental processing of linguistic input. Unlike many other formalisms, DS models the incremental building up of *interpretations* without presupposing or indeed recognising an independent level of syntactic processing. Thus, the output for any given string of words is a purely *semantic* tree representing its predicate-argument structure; words and grammatical rules correspond to actions which incrementally license the construction of such representations in tree format, employing a modal logic for tree description which provides operators able to introduce constraints on the further development of such trees (LOFT, Blackburn and Meyer-Viol, 1994). The DS lexicon consists of *lexical actions* keyed to words, and also a set of globally applicable *computational actions*, both of which constitute packages of monotonic update operations on semantic trees, and take the form of IF-THEN action-like rules which when applied yield semantically transparent structures. For example, the lexical action corresponding to the word *john* has the preconditions and update operations in example (27): if the pointer object ($\diamondsuit$), which indicates the node being checked on the tree, is currently positioned at a node that satisfies the properties of the precondition, (e.g. has the requirement type $?Ty(e)$), then all the actions in the post-condition can be completed, these being simple LOFT monotonic tree operations.

(27) IF $?Ty(e)$
THEN put($Ty(e)$)
put($\left[\ x\ :\ john\ \right]$)
ELSE abort

The trees upon which actions operate represent terms in the typed lambda calculus, with mother-daughter node relations corresponding to semantic predicate-argument structure (see (1) below). In DS-TTR, the nodes of such trees are annotated with a node type (e.g. $Ty(e)$) and semantic formulae in the form of TTR *record types* Cooper (2005). In the recent move to incorporate TTR into DS (Purver et al., 2010, 2011), following Cooper (2005), TTR *record types* consist of fields of the form $\left[\ l\ :\ T\ \right]$, containing a unique label l in the record type and a type T which represents the node type of the DS tree at which the formula is situated if it is a simple type, or else the *final* node type (e.g. type t for a predicate at a $Ty(e_s \rightarrow t)$ node). Fields can be *manifest* (i.e. have a singleton type such as $\left[\ l_{=a}\ :\ T\ \right]$). Within record types there can be *dependent* fields such as those whose singleton type is a predicate as in $\left[\ p_{=like(x,y)}\ :\ t\ \right]$, where x and y are labels in fields preceding it (i.e. are higher up in the graphical representation). Functions from record type to record type in the variant of TTR we use here employ paths, and are of the form $\lambda r : \left[\ l1\ :\ T1\ \right]$ $\left[\ l2_{=r.l1}\ :\ T1\ \right]$, an example being the formula at the type $Ty(e_s \rightarrow t)$ node in the trees in (1) below, giving DS-TTR the required functional application capability: functor node functions are applied to their sister argument node's formula, with the resulting β-reduced record type added to their mother[1].

We further adopt an event-based semantics along Davidsonian lines (Davidson, 1980). As shown below, we include an event node (of type e_s) in the representation: this allows tense and aspect to be expressed[2], allowing incremental modification to the the record type on the $Ty(e_s)$ node during parsing and generation after its initial placement in the initial axiom tree. The inclusion of an event node also permits a straightforward analysis of optional adjuncts as extensions of an existing semantic representation (see below section 4.1 and Appendix 1 for examples).

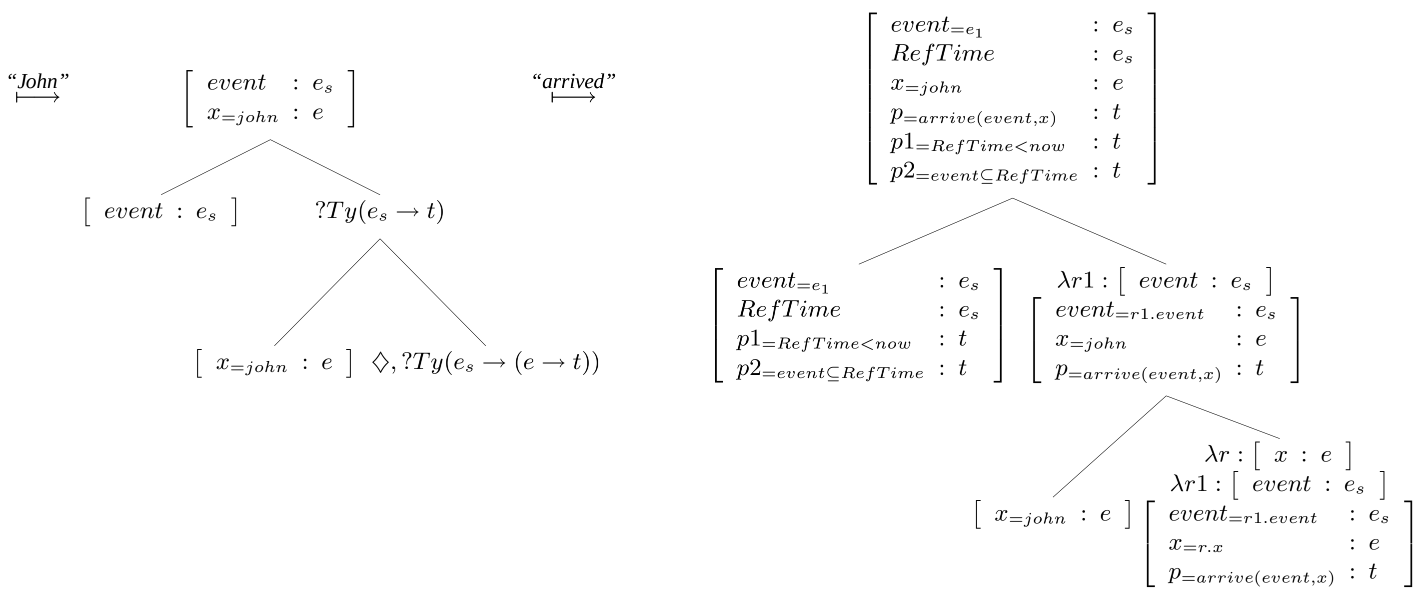

Figure 1: Parsing "John arrived"

DS-TTR parsing intersperses the testing and application of both lexical actions triggered by input words and the execution of permissible sequences of computational actions, with their updates monotonically constructing the tree. Central to this perspective is the concept of structural underspecification with subsequent update, a stance which is reflected by including among the tree-transitions to be induced, one which yields a tree relation with no more characterisation than $\langle\uparrow_*\rangle Tn(a)$; this dictates that the node so constructed should be dominated by some node in a definable tree-domain later to be updated when a suitable fixed tree-node relation becomes available. This approach, familiar in parsing implementations of long-distance dependency (see **?**, also Kaplan and Zaenen (1989) and the concept of *functional uncertainty* of LFG), is incorporated into DS as a core structural transition of an unfixed node creation and subsequent merge with the matrix tree, and is taken as the basis for a broad range of long-distance and other non-contiguous dependencies. All such cases are made subject to resolution at some future point through the imposition of a requirement for a fixed tree-node value $?\exists x Tn(x)$.

[1] For functional application and Link-Evaluation (see Cann et al. (2005b, ch. 3), but also Appendix 1 for example DS-TTR derivations involving Link-Evaluation), which require the intersection/concatenation of two record types, *relabelling* is carried out when necessary to avoid leaving incorrect variable names in the record types in the manner of Cooper (2005) and Fernández (2006).

[2] see Cann (2011) for the detailed Reichenbachian treatment of tense/aspect used here.

Seen in these terms, successful parses are sequences of action applications that lead to a tree which is complete (i.e. has no outstanding requirements on any node, and has type $Ty(t)$ at its root node as in (1)). Incomplete *partial* structures are maintained in the parse state on a word-by-word basis, giving DS its incrementality, and with the DS-TTR composite it is now possible to make available a record type which gives the maximal amount of semantic information available for partial as well as complete trees (see the left tree in (1) above) by a simple tree compiling operation, which is schematically:

1. Decorate all terminal argument nodes (the left side nodes) lacking instantiated formulae with record types containing a variable of the appropriate type.

2. Carry out functional application from the record types of compiled functor nodes to the record types of their sister argument nodes in a bottom-up fashion, compiling a β-reduced record type at their mother node. Relabel record type variables where necessary. For argument nodes with no sister functor nodes, simply *merge* them (return the *meet* type (Cooper, 2005)) with the current root node's record type.

This TTR compilation efficiently solves the problem of the previously implicit strong incremental semantic representation in DS, as now maximal record types become available as each word is processed.

Finally, in DS, as well as matrix trees, (island) structures can be induced as locally independent simple predicate-argument structures, so-called *linked* trees, which are twinned as an asymmetric non-structural tree-dependency ensured through a sharing of formula terms at nodes in the two trees in question, incrementally imposed in the transition from development of one partial tree to the other (see Kempson et al., 2001). Canonical cases are relative clause adjuncts (Cann et al., 2005b; Gregoromichelaki, 2006), but equally, the LINK transition applies to a broad range of phenomena such as adjuncts and hanging-topic constructions. Within DS-TTR, LINKs are elegantly evaluated as the intersection/concatenation (the *meet* operation, as in Cooper (2005)) of the record-type accumulated at the top of a LINKed tree and the matrix tree's root node record type (see Appendix 1 for example derivations).

The advantage of the DS-TTR composite system is the meta-theoretical clarity it affords to the growth process defined by the modular LOFT-TTR architecture. In particular, the LOFT underpinnings to the mechanisms of tree-growth mean that the DS insight that core syntactic restrictions emerge as immediate consequences of the LOFT-defined tree-growth dynamics is preserved without modification (See Cann et al. (2005b), Cann et al. (2007); Kempson and Kiaer (2010); Kempson et al. (2011a); Chatzikyriakidis and Kempson (2011)).

3.2 DS-TTR procedural context as a graph

Aside from the strong incremental interpretation that DS-TTR representations afford, in line with the stipulations for adequate models of dialogue, our model provides the incremental access to *procedural context* required for modelling the phenomena reviewed above. For DS, this context is taken as including not only the end product of parsing or generating an utterance (the semantic tree and corresponding string), but also information about the dynamics of the parsing process itself – the lexical and computational action sequence used to build the tree. As defined in Purver and Kempson (2004b); Purver et al. (2006), one possible model for such a context can be expressed in terms of triples $\langle T, W, A\rangle$ of a tree T, a word-sequence W and the sequence of actions A, both lexical and computational, that are employed to construct the trees. In parsing, the parser state P at any point is characterised as a set of these triples; in generation, the generator state G consists of a goal tree T_G and a set of possible parser states paired with their hypothesised partial strings S. As will be addressed below, the definition of a parser/generator state in terms of parse states ensures equal access to context for parsing and generation, as required, with each able to use a full representation of the dynamic linguistic context produced so far.

A further modification provides the required *incremental representation* as stipulated above. This modification requires changing the view of linguistic context as centring around a set of essentially unrelated action sequences; an alternative is to characterise DS procedural context as a Directed Acyclic Graph (DAG). Sato (2011) shows how a DAG with DS *actions* for edges and (partial) *trees* for nodes allows a compact model of the dynamic parsing process; and Purver et al. (2011) extend this to integrate it with a word hypothesis graph (or "word lattice") as obtained from a standard speech recogniser.

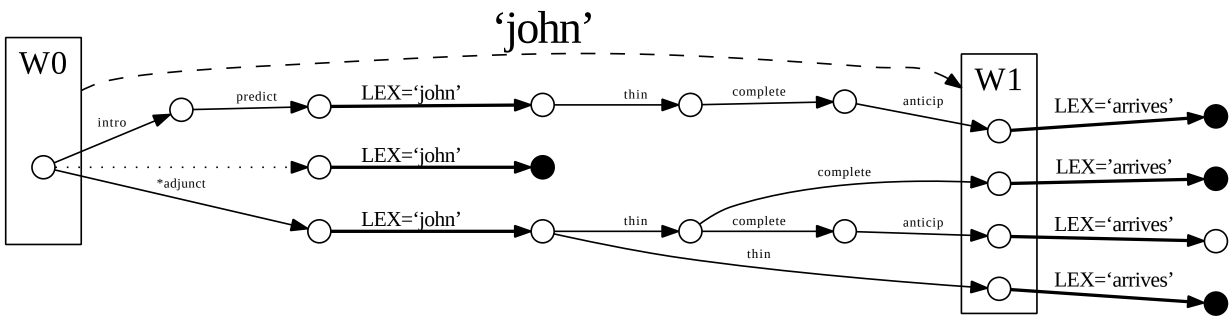

Figure 2: DS context as DAG, consisting of parse DAG (circular nodes=trees, solid edges=lexical(bold) and computational actions) *groundedIn* the corresponding word DAG (rectangular nodes=tree sets, dotted edges=word hypotheses) with word hypothesis 'john' spanning tree sets W0 and W1.

The graphical characterization results in a model of context as shown in figure 2, a hierarchical model with DAGs at two levels. At the action level, the parse graph DAG (shown in the lower half of figure 2 with solid edges and circular nodes) contains detailed information about the actions (both lexical and computational) used in the parsing or generation process: edges corresponding to these actions are connected to nodes representing the partial trees built by them, and a path through the DAG corresponds to the action sequence for any given tree. At the word level, the word hypothesis DAG (shown at the top of figure 2 with dotted edges and rectangular nodes) connects the words to these action sequences: edges in this DAG correspond to words, and nodes correspond to sets of parse DAG nodes (and therefore sets of hypothesized trees). For any partial tree, the context (the words, actions and preceding partial trees involved in producing it) is now available from the paths back to the root in the word and parse DAGs. Moreover, the sets of trees and actions associated with any word or word subsequence are now directly available as that part of the parse DAG spanned by the required word DAG edges. This, of course, means that the contribution of any word or phrase can be directly obtained, fulfilling the criterion of incremental representation. It also provides a compact and efficient representation for multiple competing hypotheses, compatible with DAG representations commonly used in interactive systems, including the incremental dialogue system Jindigo (Skantze and Hjalmarsson, 2010), a move that has been taken by Purver et al. (2011). Importantly, as described below, the DS definition of generation in terms of parsing still means this model will be equally available to both, and used in the same way by both modules. The criteria of interchangeability and equal incremental access to context, essential for the modelling of covering compound contributions and self-repairs (see below), are therefore satisfied.

3.3 DS-TTR Generation as Parsing

In turning to the DS-TTR account of generation, a number of preliminaries have first to be addressed. First, as the split utterance data demonstrate, incremental behaviour needs to include allowing confirmation behaviour as in (23), (20), continuations in utterances shared between the user and system as in (3), (4), (7)-(23), but also user interruptions without discarding the semantic content built up so far to provide for realistic clarification and *self-repair* capability such as in (2). And as these data have illustrated, individual fragments may display more than one such attribute. As we have stipulated above, the three requirements of exhibiting *strong incremental interpretation*, *incremental representation* on a word-by-word basis and continual access to *procedural context* extends to the generation module, which must implement all information made available by selected expressions without delay. There is however also a fourth requirement in generation: the generation of incremental dialogue phenomena of course requires incremental parsers and dialogue management modules which can reason with the semantic representations it produces, so an extrinsic necessity on the module is that it should have the property of *representational interchangeability* with other modules. DS-TTR can meet these criteria, which conventional grammar frameworks, as we have already seen, struggle to capture elegantly, particularly for examples such as (10) and (11).

Amongst recent developments in incremental generation, Guhe (2007) models incrementality in the conceptualization phase, developing a module which generates semantic input to the formulator incrementally. While syntactic formulation is not the focus, the interface between the incremental conceptualizer and the formulator is clearly defined: the conceptualizer's incremental modification to pre-verbal messages characterizes downstream tactical generation and the modification of the messages with correction increments causes self-repair surface forms to be realized. Buß and Schlangen (2011) recently introduced dialogue management strategies in the same spirit and albeit less psychologically motivated, Skantze and Hjalmarsson (2010) provide a similar approach to Guhe's conceptual change model in their implementation of incremental speech generation in a dialogue system. Generation input is defined in terms of canned-text *speech plans* sent from the dialogue manager that are divided up into word-length *speech units*. The procedure consists of the incremental vocalization of each unit, coupled with self-monitoring the plan in the sense of Levelt (1989). As speech plans may change dynamically during interaction with a user, upon detection of difference by the monitor through a simple string-based comparison of the incoming plan with the current one, both *covert* and *overt* self-repairs can be generated, depending on the number of units in the plan realized at the point of detection. These approaches thus variously utilise a system of partial inputs to generation components to reduce complexity burdens and top-down revision of string-based speech plans or syntactic structures, however there is not a clear description of how an incremental semantic representation can be tightly coupled with surface realisation to facilitate fine-grained build up of meaning during generation, which is a prerequisite for generating interesting incremental dialogue phenomena. Skantze and Hjalmarsson's model is a step towards coupling word-by-word generation and self-monitoring, however the lack of incremental semantics and domain-general grammar makes scalability and integration with a parsing module difficult. Relating semantics to surface form via canned text restricts the system's possible utterances hugely even in one domain. And in an account such as this, based on full sentence characterisation with late deletion, there is no semantic word-by-word incrementality to the form of explanation, so dynamic ongoing alteration is precluded in principle.

In comparison to these, the DS system addresses the incrementality problem head on by incorporating within the grammar formalism at least some of the necessary incrementality requirements for dialogue. And this is extended to generation in a very direct way as Purver and Kempson (2004a) demonstrate. An incremental DS model of surface realisation can be neatly defined in terms of the DS parsing process and a *subsumption check* against a *goal tree*. The DS generation process is word-by-word incremental with maximal tree representations continually available, and it effectively combines lexical selection and linearisation into a single action due to the word-by-word iteration through the lexicon. Also, while no formal model of self-repair has hitherto been proposed in DS (but see below section 4.2), self-monitoring is inherently part of the generation process, as each word generated is parsed. However, while the Purver and Kempson (2004a) DS generation model is incremental, it does not meet the criterion of strict incremental interpretation as stipulated above, as maximal information about the dependencies between the semantic formulae in the tree may not be computed until the tree is complete - this is an issue addressed in the developments reported here. Also, in terms of logical input forms to generation, the goal tree needs to be constructed from the grammar's actions, so any dialogue management module must have full knowledge of the DS parsing mechanism and lexicon, and so interchangeability of representation becomes difficult. For this reason several adjustments are suggested below, given the new DS-TTR framework.

3.3.1 TTR goal concepts and subtype checking for lexicalisation

One straightforward modification to the DS generation model enabling representational interchangeability with other modules is the replacement of the previously defined *goal tree* with a *TTR goal concept*, which takes the form of a record type such as:

(28)
$$\left[\begin{array}{lcl} event_{=e1} & : & e_s \\ RefTime_{=today} & : & e_s \\ p1_{=RefTime\bigcirc event} & : & t \\ x1_{=Sweden} & : & e \\ p2_{=from(event,x1)} & : & t \\ x_{=robin} & : & e \\ p_{=arrive(event,x)} & : & t \end{array}\right]$$

Importantly, the goal concept may be *partial*, in that the dialogue manager may further specify it, and it need not correspond to a complete sentence, which is important for incremental dialogue management strategies (Guhe, 2007; Buß and Schlangen, 2011), as it is needed for such examples as (1)-(3). This move also means the dialogue manager may input goal concepts directly to the generator, and no considerations of the requirements of the DS grammar are needed, in contrast to Purver and Kempson (2004a)'s approach. The tree subsumption check in the original DS generation model can now be characterized again as a TTR subtype relation check between the goal tree and the trees in the parse state's compiled TTR formulae:

(29) *Subtype relation check*
For record types $p1$ and $p2$, $p1 \sqsubseteq p2$ holds just in case for each field $\left[\ l\ :\ T2\ \right]$ in $p2$ there is a field $\left[\ l\ :\ T1\ \right]$ in $p1$ such that $T1 \sqsubseteq T2$, that is to say just in case any object of type $T1$ is also of type $T2$.[3] The type inclusion relation is reflexive and transitive. (adapted from Fernández (2006, p.96))

An example of a successful generation path is shown in Figure 3[4], where the incremental generation of "john arrives" succeeds as the successful lexical action applications at transitions $\boxed{1}\mapsto\boxed{2}$ and $\boxed{3}\mapsto\boxed{4}$ are interspersed with applicable computational action sequences at transitions $\boxed{0}\mapsto\boxed{1}$ and $\boxed{2}\mapsto\boxed{3}$, at each stage passing the subtype relation check with the goal (i.e. the goal is a subtype of the top node's compiled record type), until arriving at a tree that *type matches* the assigned goal concept in $\boxed{4}$ in the rich TTR sense of *type*. In implementational terms, there will in fact be multiple generation paths in the generation state, including incomplete and abandoned paths, which can be incoporated into the DS notion of context as a DAG.

Another advantage of working with TTR record types rather than trees during generation is that selecting relevant lexical actions from the lexicon can take place before generation begins through comparing the semantic formulae of the actions to the goal concept. Subtype checking makes it possible to reduce the computational complexity of lexical search through a pre-verbal lexical action selection. Informally, a sublexicon $SubLex$ can be created when the goal concept $GoalTTR$ is inputted to the generator by the following process:

(30) *Pre-verbal lexicalisation*
For all lexical actions L_i in the lexicon, add to $SubLex$ if $GoalTTR$ is a subtype of the TTR record type or range of the TTR record type function added by L_i.

[3] Importantly, this also holds in the case of *manifest* types, as while the notation $\left[\ l_{=v}\ :\ T2\ \right]$ is used in this paper, this is syntactic sugar for $\left[\ l\ :\ T2_v\ \right]$, so in these cases for $p1 \sqsubseteq p2$ to hold, for each field $\left[\ l\ :\ T2_v\ \right]$ in $p2$ there is a field $\left[\ l\ :\ T1_v\ \right]$ in $p1$ such that $T1_v \sqsubseteq T2_v$

[4] Since Figure 3 is given to display the generation path dynamics, event term specifications are omitted for simplicity.

Depending on a system designer or experimenter's choice of how many fields a DS-TTR lexical action's TTR formulae has, the size of $SubLex$ will vary. For instance if lexical actions for verbs lack a field for tense information, several candidates may be selected in $SubLex$ which are all valid supertypes of the goal concept (e.g. likes, like, liked), and less appropriate candidates may be filtered out at a later stage. With this move, the more lexicalised the grammar, the smaller $SubLex$ will be, and consequently the smaller the search space for generation. It is also worth noting that semantically underspecified lexical entries, such as those for 'do'-type auxiliaries used in verb phrase ellipsis, may be selected here by default, as the values in their fields are null and inherit values from context (Kempson et al., 2011b), so anaphoric and elliptical forms are readily available.

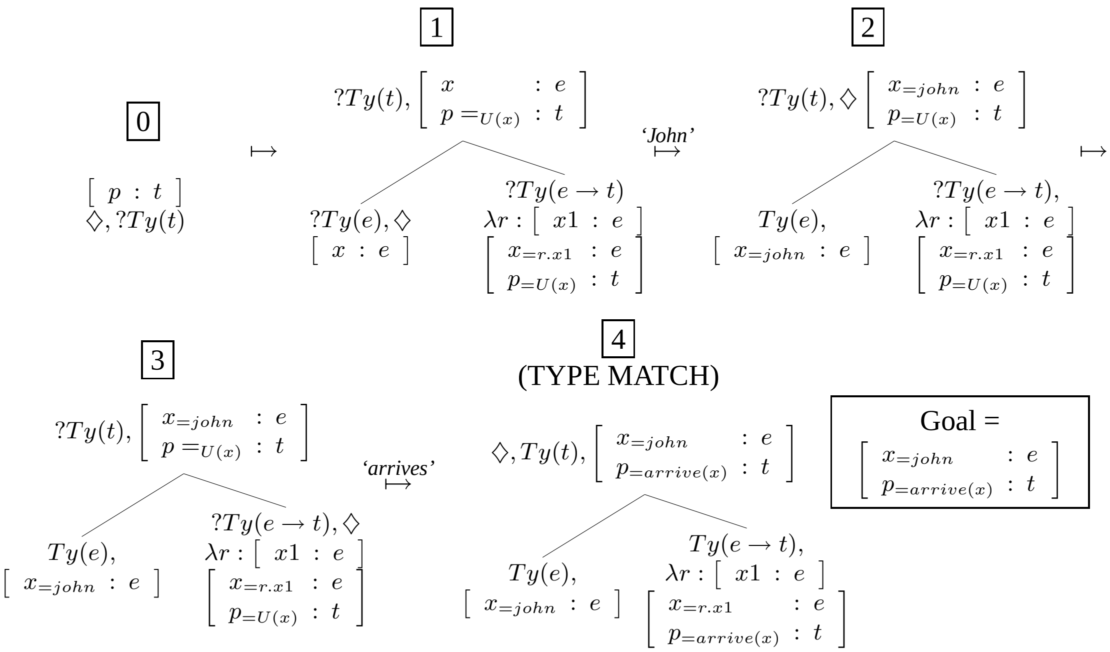

Figure 3: Successful generation path in DS-TTR

3.4 Implementation: DyLan dialogue system

`DyLan` Eshghi et al. (2011), a prototype dialogue system utilising the DS-TTR implementation in parsing and generation, has been implemented in Java[5] within the incremental dialogue system framework Jindigo (Skantze and Hjalmarsson, 2010), utilising the incremental unit (IU) graphs in the system module's input and output buffers based on Schlangen and Skantze (2009)'s IU model. Following Sato's (2011) insight that the procedural context of DS parsing can be characterized in terms of graphical search as described in section 3.2 and following Purver et al.'s (2011) implementation, the parse state of the parsing module is characterized as three linked directed acyclic graphs (DAGs): (1) a linearly constructed (no backtracking allowed) word hypothesis graph, consisting of word hypothesis edge IUs between vertices W_n, which have *groundedIn* links (i.e. dependency relations) to edges in (2) the DS parsing graph, which adds parse state edge IUs between vertices S_n (whose internal state is a DS tree), which in turn have groundedIn relations to edges in (3) the concept graph which has domain concepts as its IUs built between vertices C_n.

In generation, the architecture is the inverse of interpretation in virtue of there being a goal concept: (1) the concept graph produces goal concepts and adds them as IU edges between vertices GC_n, (2) the DS parsing graph is incrementally constructed on a word-by-word basis by testing the lexical actions in the

[5]Available from http://dylan.sourceforge.net/

sublexicon produced for current goal concept (see section 3.3), and (3) the word graph's edges are `add`ed to the output buffer of the module during word-by-word generation, but only `commit`ted (made available to the downstream vocalizer) when they lead to trees whose TTR formulae for which the current goal concept is a valid subtype (i.e. they form part of a valid generation path as in figure 3).

4 Incremental processing of dialogue phenomena

By way of explanation of the dialogue phenomena, we can now see how the overall `DyLan` dialogue system deals with them in parsing and generation, using the mechanisms of DS-TTR as set out above.

4.1 Compound contributions

Previous formal and computational accounts of compound contributions (CCs) have focussed on *completions* in which, by definition, a responder succeeds in projecting a string the initial speaker had intended to convey. The foremost implementation is that of Poesio and Rieser (2010), using the PTT model for incremental dialogue interpretation (Poesio and Traum, 1997; Poesio and Rieser, 2003) in combination with LTAG (Demberg and Keller, 2008). The approach is grammar-based, incorporating syntactic, semantic and pragmatic information via the lexicalised TAG grammar paired with their PTT model, providing an account of the incremental interpretation process, incorporating lexical, syntactic and semantic information. Beyond this, they provide a detailed account of how a suggested collaborative completion might be derived using inferential processes and the recognition of plans: by matching the partial representation at speaker transition against a repository of known plans in the relevant domain, an agent can determine the components of these plans which have not yet been made explicit and make a plan to generate them. This model therefore meets many of the criteria defined above: both interpretation and representation are incremental, with semantic and syntactic information being present; the use of PTT suggests that linguistic context can be incorporated suitably. However, while reversibility might be incorporated by choice of suitable parsing and generation frameworks, this is not made explicit; and the extensibility of the representations seems limited by TAG's approach to adjunction (extension via syntactic adjuncts seems easy to treat in this approach, but more general extension is less clear). The use of TAG also seems to restrict the grammar to licensing grammatical *strings*, problematic for some CCs (e.g. examples (10) and (11) above, in which *semantic* dependencies hold between the two parts of the CC); and the mechanism may not be sustainable for the broad range of data where the participants make no attempt to match what the other party might have in mind. Moreover, as with other syntactic accounts, whenever such mechanism is used, this will lead directly to predictions of processing complexity that we have strong reason to believe will not be met.

In the `DyLan` model, a broad range of compound utterances now follows as an immediate consequence of DS-TTR. The use of TTR record types removes the need for grammar-specific parameters; and the interchangeability of representations between parsing and generation means that the construction of a data structure can become a collaborative process between dialogue participants, permitting a range of varied user input behaviour and flexible system responses. This use of the same representations by parsing and generation guarantees the ability to begin parsing from the end-point of any generation process, even mid-utterance; and to begin generation from the end-point of any parsing process: the successive sequential exchanges between participants leading to a collaboratively completed utterance is directly predicted, as in (8), (9) and elsewhere. Both parsing and generation models are now characterised entirely by the parse context DAG with the addition for generation of a TTR goal concept. The transition from generation to parsing becomes almost trivial: the parsing process can continue from the final node(s) of the generation DAG, with parsing actions extending the trees available in the final node set as normal. Transition from parsing to generation also requires no change of representation with the DAG produced by parsing acting as the initial structure for generation (figure 4) though we require the addition of a goal concept to drive the generation process. Given the incremental interpretation provided by the use of record types throughout, we can now also see how a generator might produce such a goal at speaker transition:

Figure 4: Completion of a compound contribution using incremental DS-TTR record type construction with parser and generator *sharing* a parse state.

The same record types are thus used throughout the system: as the concepts for generating system plans, as the goal concepts in NLG, and for matching user input against known concepts in suggesting continuations. Possible system transition points trigger alternation between modules in their co-construction of the shared parse/generator; in `DyLan` this is provided by a simplistic dialogue manager with high-level methods without reference to syntax or lexical semantics. A goal concept can be produced by the dialogue manager at a speaker transition by searching its domain concepts for a suitable subtype of the TTR record type built so far, guaranteeing a grammatical continuation given the presence of appropriate lexical actions and allowing exchanges such as (1). This extends the method for compound contributions described in Purver and Kempson (2004a), however now the dialogue manager has an elegant decision mechanism for aiding content selection. And, given the presumption of context, content and goal specifications all in terms of record types, the ability to construct goals in a scenario without linguistic antecedents is also allowed for (12), (13) and (15) above.

The data of compound contributions thus follows in full, even when either the goal record type for the interrupter does not match that of the initiator as in (5), or when the goal record type does not correspond to a complete domain concept, as in the successive fragment exchanges such as (9).This is achieved through progressive extensions of the partial tree so far, either directly, or by adding LINKed trees as required for adjunctive phenomena. This results in the word-by-word further specification of the record type at the root of the matrix tree representing the maximal interpretation of the string/utterance so far. In Figure 5 we give the progressive record-type specification for the exchange (31), a simplification of (9), showing how incomplete structures may serve as both input and output for either party:

(31) A: Today Robin arrives

B: from

A: Sweden

Details of the tree derivations are omitted in Figure 5, but we have included these in Appendix 1, which contains a fuller tree derivation for (31) plus an ‘other-correction’ as modelled identically to self-repair as set out in the next section.

$$\left[\begin{array}{lcl} event & : & e_s \\ RefTime_{=today} & : & e_s \\ p & : & t \end{array}\right] \mapsto \left[\begin{array}{lcl} event_{=e1} & : & e_s \\ RefTime_{=today} & : & e_s \\ p1_{=RefTime \bigcirc event} & : & t \\ x_{=robin} & : & e \\ p_{=arrive(event,x)} & : & t \end{array}\right] \mapsto \left[\begin{array}{lcl} event_{=e1} & : & e_s \\ RefTime_{=today} & : & e_s \\ p1_{=RefTime \bigcirc event} & : & t \\ x_{=robin} & : & e \\ p_{=arrive(event,x)} & : & t \\ x1 & : & e \\ p2_{=from(event,x1)} & : & t \end{array}\right] \mapsto \left[\begin{array}{lcl} event_{=e1} & : & e_s \\ RefTime_{=today} & : & e_s \\ p1_{=RefTime \bigcirc event} & : & t \\ x_{=robin} & : & e \\ p_{=arrive(event,x)} & : & t \\ x1_{=Sweden} & : & e \\ p2_{=from(event,x1)} & : & t \end{array}\right]$$

“A: Today” $\mapsto$ *“..Robin arrives”* $\mapsto$ *“B: from?”* $\mapsto$ *“A: Sweden”*

Figure 5: Incremental interpretation via TTR subtypes

As noted, more complex forms can be generated by incorporating LINKed trees, as is presumed in the characterisation of the many extensions by the addition of an adjunct, as in (8), (11), (18) (See Appendix 1), without any of these having to involve any extension of the formal DS vocabulary.

4.2 Self-repair

In this section, we present our initial model of self-repair. In generation, as a goal concept may be revised shortly after or during the generation process due to a decision by the dialogue manager, trouble in generating the next word may be encountered. `DyLan`'s repair function operates if there is an empty state, or no possible DAG extension, after the semantic filtering stage of generation (resulting in no candidate succeeding word edge) by restarting the generation procedure from the last committed parse state edge. It continues backtracking by one vertex at a time in an attempt to extend the DS DAG until successful, as can be seen in figure 6. Note that the previously committed word graph edge for *London* is not revoked, following the principle that it has been in the public record and hence should, correctly, still be accessible. Clark (1996) makes this point about utterances such as "the interview was.. it was alright" where the *reparandum* (repaired material) still needs to be accessed for the anaphoric use of *it* to succeed.

Our protocol is consistent with Shriberg and Stolcke (1998)'s empirical observation that the probability of retracing N words back in an utterance is more likely than retracing from N+1 words back, making the repair as local as possible. Utterances such as "I go, uhh, leave from Paris" are generated incrementally, as the repair is integrated with the semantics of the part of the utterance before the repair point, maximising re-use of existing semantic structure, while the time-linear word graph continues to extend but with the repair's edges *groundedIn* different paths of the parse DAG to the reparandum's edges (as in Fig.6; see also (2)).

A subset of self-repairs, *extensions*, where the repair effects an "after-thought", usually in transition relevant places in dialogue after apparently complete turns, is dealt with straightforwardly by our module: e.g. (8), (1)-(3), (9), (18). The DS parser treats these as monotonic growth of the matrix tree through LINK adjunction (Cann et al., 2005b), resulting in subtype extension of the root TTR record type. Thus, a change in goal concept during generation will not always put demands on the system to backtrack, such as in generating the fragment after the pause in "I go to Paris . . . from London". It is only at a semantics-syntax mismatch where the revised goal TTR record type does not correspond to a permissible extension of a DS tree in the DAG as in Fig.6, where overt repair will occur.

Figure 6: Incremental DS-TTR generation of a self-repair upon change of goal concept. Type-matched record types are double-circled nodes and revoked edges indicating failed paths are dotted. Inter-graph *groundedIn* links go from top to bottom.

Note that the mechanism for recovery of meaning in parsing a self-repaired utterance can be defined in a similarly local way in our model, using the following definition:

(32) **Repair** IF from parsing word W there is no edge SE_n able to be constructed from vertex S_n (no parse) or if no domain concept hypothesis can be made through subtype relation checking, `repair`: parse word W from vertex S_{n-1} and and should that parse be successful add a new edge to the top path, without removing any committed edges beginning at S_{n-1}.

It is worth noting that in contrast to Skantze and Hjalmarsson's (2010) string-based *speech plan* comparison approach, there is no need to regenerate a fully-formed string from a revised goal concept and compare it with the string generated thus far to characterize repair. Instead, repair is driven by attempting to extend existing parse paths to construct the new target record type, *retaining* the semantic representation and the procedural context of actions already built up in the generation process to avoid the computational demand of constructing syntactic structures from afresh where possible.

4.3 Speech Acts and speaker/hearer attributions in DS/TTR

A further bonus of combining DS mechanisms with TTR record types as output decorations is the allowance of a much richer vocabulary for such decorations, as empirically warranted. In particular, it provides a basis from which speaker and hearer attributes may be optionally specified. In this connection, Purver et al. (2010) propose a specification of fields with sub-field specifications, one a $contxt$ sub-field for speaker-hearer values, the second, $contnt$, for familiar lambda-terms, a modification which allows a record of speaker-hearer attributions to be optionally kept alongside function-argument content record type specifications so that the different anaphor-dependency resolutions across switch of participant roles can be modelled as in (10)-(11) without disturbing content compilation of the lambda terms. No details are given here (see Purver et al. (2010) for details); but in principle with unification of record types available for record types of arbitrary complexity, such specifications are unproblematic. The optionality of specification of speaker/hearer relations/attributes raises issues of what constitutes successful communication, in particular for Gricean and proto-Gricean models in which recognition of the content of the speaker's intentions is essential: Poesio and Rieser (2010) is illustrative. We do not enter into this debate here, but merely note that this stance is commensurate with the data of section 1 in which participants' intentions may emerge or be subject to modification during the course of a conversation without jeopardising its success (see Gregoromichelaki et al. (2011); **?** for detailed discussion).

5 Conclusion

We have presented a formal framework for modelling conversational dialogue with parsing and generation modules as controlled by a dialogue manager, both of which reflect word by word incrementality, using a hybrid of Dynamic Syntax and Type Theory with Records. The composite framework allows access to record types incrementally during generation, providing strict incremental representation and interpretation for substrings of utterances that can be accessed by existing dialogue managers, parsers and generators equally, allowing the articulation of syntactic and semantic dependencies across parser and generator modules. Characterising DS generation as a DAG in tandem with a DAG-based parser, in particular, allows easy integration into incremental dialogue systems, and facilitates goal revision and self-repairing capabilities. Retaining the DS assumption of tree growth as defined in LOFT as input to both parsing and generation systems preserves the original expressibility of syntactic generalisations unaltered. The model also allows for experimentation with search techniques, which will be explored in coming work. The account of quantification of the earlier DS system Kempson et al. (2001) depended on the lower type account of quantification as expressed through epsilon terms definable in the epsilon calculus. This system, though equivalent in expressive power to classical predicate logic, and hence relatively restricted given natural language expressivity, is nonetheless not incommensurable with the more general type-dependent account of quantification (see Fernando (2002), Cooper (2012)) made available by the Martin-Löf type-logical proof system. With the work on developing the DS-TTR composite framework having reached current levels of formal explicitness, work on exploring mappings from the DS model of quantification onto TTR accounts of quantificational dependency that preserve the incrementality of scope dependency choice made available in that earlier DS account thus now becomes the next important challenge on the horizon.

References

James Allen, George Ferguson, and Amanda Stent. An architecture for more realistic conversational systems. In *Proceedings of the 2001 International Conference on Intelligent User Interfaces (IUI)*, January 2001.

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

Robert B. Brandom. *Making it Explicit: Reasoning, Representing, and Discursive Commitment*. Harvard University Press, 1994.

Okko Buß and David Schlangen. Dium : An incremental dialogue manager that can produce self-corrections. In *Proceedings of SemDial 2011 (Los Angelogue), Los Angeles, CA*, pages 47–54, 2011.

Ronnie Cann. Towards an account of the english auxiliary system: building interpretations incrementally. In Ruth Kempson, Eleni Gregoromichelaki, and Christine Howes, editors, *Dynamics of Lexical Interfaces*. Chicago: CSLI Press, 2011.

Ronnie Cann, Tami Kaplan, and Ruth Kempson. Data at the grammar-pragmatics interface: the case of resumptive pronouns in English. *Lingua*, 115(11):1475–1665, 2005a. Special Issue: On the Nature of Linguistic Data.

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

Ronnie Cann, Ruth Kempson, and Matthew Purver. Context and well-formedness: the dynamics of ellipsis. *Research on Language and Computation*, 5(3):333–358, 2007.

S. Carberry. *Plan recognition in natural language dialogue*. the MIT Press, 1990.

Stergios Chatzikyriakidis and Ruth Kempson. Standard modern and pontic greek person restrictions: A feature-free dynamic account. *Journal of Greek Lingusitics*, pages 127–166, 2011.

N. Chomsky. *Aspects of the Theory of Syntax.* MIT, Cambridge, MA, 1965.

Noam Chomsky. *New Horizons in the Study of Language and Mind*. Cambridge University Press, 2000.

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

Herbert H. Clark and Jean E. Fox Tree. Using *uh* and *um* in spontaneous speaking. *Cognition*, 84(1):73–111, 2002.

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

Robin Cooper. Type theory and semantics in flux. In Ruth Kempson, Nicholas Asher, and Tim Fernando, editors, *Handbook of the Philosophy of Science*, volume 14: Philosophy of Linguistics, pages 271–323. North Holland, 2012.

Donald Davidson. Truth and meaning. *Synthese*, 17:304–323, 1967.

Donald Davidson. *Essays on Actions and Events*. Clarendon Press, Oxford, UK, 1980.

Donald Davidson. A nice derangement of epitaphs. In E. Lepore, editor, *Truth and Interpretation*, pages 433–446. 1986.

Donald Davidson. The social aspect of language. In B. McGuiness and G. Oliveri, editors, *The philosophy of Michael Dummet*. Kluwer Academic Publishers, 1994.

V. Demberg and F. Keller. A psycholinguistically motivated version of tag. In *Proceedings of the International Workshop on Tree Adjoining Grammars*, 2008.

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

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

Tim Fernando. Three processes in natural language interpretation. In W. Sieg, R. Sommer, and C. Talcott, editors, *Reflections on the Foundations of Mathematics: Essays in Honor of Solomon Feferman*, pages 208–227. Association for Symbolic Logic, Natick, Mass., 2002.

Jonathan Ginzburg. *The Interactive Stance: Meaning for Conversations*. Oxford University Press, 2012.

Jonathan Ginzburg and Robin Cooper. Clarification, ellipsis, and the nature of contextual updates in dialogue. *Linguistics and Philosophy*, 27(3):297–365, 2004.

C. Goodwin. *Conversational organization: Interaction between speakers and hearers*. Academic Press, New York, 1981.

E. Gregoromichelaki. *Conditionals: A Dynamic Syntax Account*. PhD thesis, King's College London, 2006.

E. Gregoromichelaki, R. Cann, and R. Kempson. On coordination in dialogue: subsentential talk and its implications. In Laurence Goldstein, editor, *On Brevity*. OUP, forthcoming.

Eleni Gregoromichelaki, Ruth Kempson, Matthew Purver, Greg J. Mills, Ronnie Cann, Wilfried Meyer-Viol, and Pat G. T. Healey. Incrementality and intention-recognition in utterance processing. *Dialogue and Discourse*, 2(1):199–233, 2011.

Barbara J. Grosz and Candace L. Sidner. Attention, intentions, and the structure of discourse. *Computational Linguistics*, 12(3):175–204, 1986.

Markus Guhe. *Incremental Conceptualization for Language Production*. NJ: Lawrence Erlbaum Associates, 2007.

J. Hawkins. *Efficiency and Complexity in Grammars*. Cambridge University Press, Cambridge, 2004.

Ronald Kaplan and Annie Zaenen. Long-distance dependencies, constituent structure, and functional uncertainty. In M. Baltin and A. Kroch, editors, *Alternative Conceptions of Phrase Structure*, pages 17–42. University of Chicago Press, Chicago, Illinois, 1989.

R. Kempson and J. Kiaer. Multiple long-distance scrambling: Syntax as reflections of processing. *Journal of Linguistics*, 46(01):127–192, 2010.

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

Ruth Kempson, Eleni Gregoromichelaki, and Christine Howes, editors. *The Dynamics of Lexical Interfaces*. CSLI - Studies in Constraint Based Lexicalism, 2011a.

Ruth Kempson, Eleni Gregoromichelaki, Wifried Meyer-Viol, Matthew Purver, Graham White, and Ronnie Cann. Natural-language syntax as procedures for interpretation: the dynamics of ellipsis construal. In A. Lecomte and S. Tronçon, editors, *Ludics, Dialogue and Interaction*, number 6505 in Lecture Notes in Computer Science, pages 114–133. Springer-Verlag, Berlin/Heidelberg, 2011b. ISBN 978-3-642-19210-4. URL `http://www.eecs.qmul.ac.uk/ mpurver/papers/kempson-et-al11ludics.pdf`.

R. Larson and G. Segal. *Knowledge of Meaning: An Introduction to Semantic Theory*. The MIT Press, 1995.

Staffan Larsson. The ttr perceptron: Dynamic perceptual meanings and semantic coordination. In *Proceedings of the 15th Workshop on the Semantics and Pragmatics of Dialogue (SemDial 2011 - Los Angelogue)*, pages 140–148, September 2011.

W.J.M. Levelt. Monitoring and self-repair in speech. *Cognition*, 14(1):41–104, 1983.

W.J.M. Levelt. *Speaking: From Intention to Articulation*. MIT Press, 1989.

S. C. Levinson. *Pragmatics*. Cambridge University Press, 2002.

Vincenzo Lombardo and Patrick Sturt. Incrementality and lexicalism: a treebank study. In S. Stevenson and P. Merlo, editors, *The Lexical Basis of Sentence Processing*, pages 137–156. John Benjamins, 2002.

John McDowell. *Meaning, Knowledge and Reality*. Harvard University Press, 1998.

J. Merchant. Fragments and ellipsis. *Linguistics and Philosophy*, 27:661–738, 2004.

Ruth Millikan. *The Varieties of Meaning: The Jean-Nicod Lectures*. MIT Press, 2004.

G. Mills and E. Gregoromichelaki. Establishing coherence in dialogue: sequentiality, intentions and negotiation. In *Proceedings of SemDial (PozDial)*, 2010.

David Milward. *Axiomatic Grammar, Non-Constituent Coordination and Incremental Interpretation*. PhD thesis, University of Cambridge, 1991.

R. Montague. Universal grammar. *Theoria*, 36:373–398, 1970.

J. Morgan. Sentence fragments and the notion sentence. In *Issues in linguistics: Papers in honor of Henry and Renée Kahane*, pages 719–751. 1973.

J.L. Morgan. Some interactions of syntax and pragmatics. In Peter Cole and Jerry L. Morgan, editors, *Syntax and Semantics, Volume 3: Speech Acts*, volume 3, pages 289–303. Academic Press, 1975.

Colin Phillips. Linear order and constituency. *Linguistic Inquiry*, 34:37–90, 2003.

Martin Pickering and Simon Garrod. Toward a mechanistic psychology of dialogue. *Behavioral and Brain Sciences*, 27:169–226, 2004.

Paul Piwek. Dialogue structure and logical expressivism. *Synthese*, 183:33–58, 2011.

Massimo Poesio and Hannes Rieser. Coordination in a PTT approach to dialogue. In *Proceedings of the 7th Workshop on the Semantics and Pragmatics of Dialogue (SEMDIAL)*, Saarbrücken, Germany, September 2003.

Massimo Poesio and Hannes Rieser. Completions, coordination, and alignment in dialogue. *Dialogue and Discourse*, 1:1–89, 2010. URL `http://elanguage.net/journals/index.php/dad/article/view/91/512`.

Massimo Poesio and David Traum. Conversational actions and discourse situations. *Computational Intelligence*, 13(3), 1997.

Matthew Purver. CLARIE: Handling clarification requests in a dialogue system. *Research on Language and Computation*, 4(2-3):259–288, 2006.

Matthew Purver and Ruth Kempson. Incremental context-based generation for dialogue. In A. Belz, R. Evans, and P. Piwek, editors, *Proceedings of the 3rd International Conference on Natural Language Generation (INLG04)*, number 3123 in Lecture Notes in Artifical Intelligence, pages 151–160, Brockenhurst, UK, July 2004a. Springer.

Matthew Purver and Ruth Kempson. Incrementality, alignment and shared utterances. In J. Ginzburg and E. Vallduví, editors, *Proceedings of the 8th Workshop on the Semantics and Pragmatics of Dialogue (SEMDIAL)*, pages 85–92, Barcelona, Spain, July 2004b.

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.

Matthew Purver, Christine Howes, Eleni Gregoromichelaki, and Patrick G. T. Healey. Split utterances in dialogue: A corpus study. In *Proceedings of the 10th Annual SIGDIAL Meeting on Discourse and Dialogue (SIGDIAL 2009 Conference)*, pages 262–271, London, UK, September 2009. Association for Computational Linguistics. URL `http://www.dcs.qmul.ac.uk/ mpurver/papers/purver-et-al09sigdial-corpus.pdf`.

Matthew Purver, Eleni Gregoromichelaki, Wilfried Meyer-Viol, and Ronnie Cann. Splitting the 'I's and crossing the 'You's: Context, speech acts and grammar. In P. Łupkowski and M. Purver, editors, *Aspects of Semantics and Pragmatics of Dialogue. SemDial 2010, 14th Workshop on the Semantics and Pragmatics of Dialogue*, pages 43–50, Poznań, June 2010. Polish Society for Cognitive Science.

Matthew Purver, Arash Eshghi, and Julian Hough. Incremental semantic construction in a dialogue system. In J. Bos and S. Pulman, editors, *Proceedings of the 9th International Conference on Computational Semantics*, pages 365–369, Oxford, UK, January 2011.

Yo Sato. Local ambiguity, search strategies and parsing in Dynamic Syntax. In E. Gregoromichelaki, R. Kempson, and C. Howes, editors, *The Dynamics of Lexical Interfaces*. CSLI Publications, 2011.

David Schlangen. *A Coherence-Based Approach to the Interpretation of Non-Sentential Utterances in Dialogue*. PhD thesis, University of Edinburgh, 2003.

David Schlangen and Gabriel Skantze. A general, abstract model of incremental dialogue processing. In *Proceedings of the 12th Conference of the European Chapter of the ACL (EACL 2009)*, pages 710–718, Athens, Greece, March 2009. Association for Computational Linguistics. URL `http://www.aclweb.org/anthology/E09-1081`.

Elizabeth Shriberg and Andreas Stolcke. How far do speakers back up in repairs? A quantitative model. In *Proceedings of the International Conference on Spoken Language Processing*, pages 2183–2186, 1998.

Gabriel Skantze and Anna Hjalmarsson. Towards incremental speech generation in dialogue systems. In *Proceedings of the SIGDIAL 2010 Conference*, pages 1–8, Tokyo, Japan, September 2010. Association for Computational Linguistics. URL `http://www.sigdial.org/workshops/workshop11/proc/pdf/SIGDIAL01.pdf`.

Dan Sperber and Deirdre Wilson. *Relevance: Communication and Cognition*. Blackwell, second edition, 1995.

R. Stainton. *Words and Thoughts: Subsentences, Ellipsis, and the Philosophy of Language*. Oxford University Press, 2006.

6 Appendix

This appendix provides a derivation for a split dialogue in which both input and output of intermediate generation and parsing steps involve partial structures, with a final step of correction:

(33) A: Today Robin arrives

B: from

A: Sweden

B: with Elizabet?

A: no, Staffan.

Notice how the event node on the matrix tree is represented $\boxed{\text{EVENT}}$ and then through expansion/modification of its type specification as successively EVENT^1, EVENT^2, etc so as to indicate its location on the tree during the build up of the other trees through LINK adjunction to it. The matrix tree type specification is not repeatedly shown here across these various revisions for reasons of space.

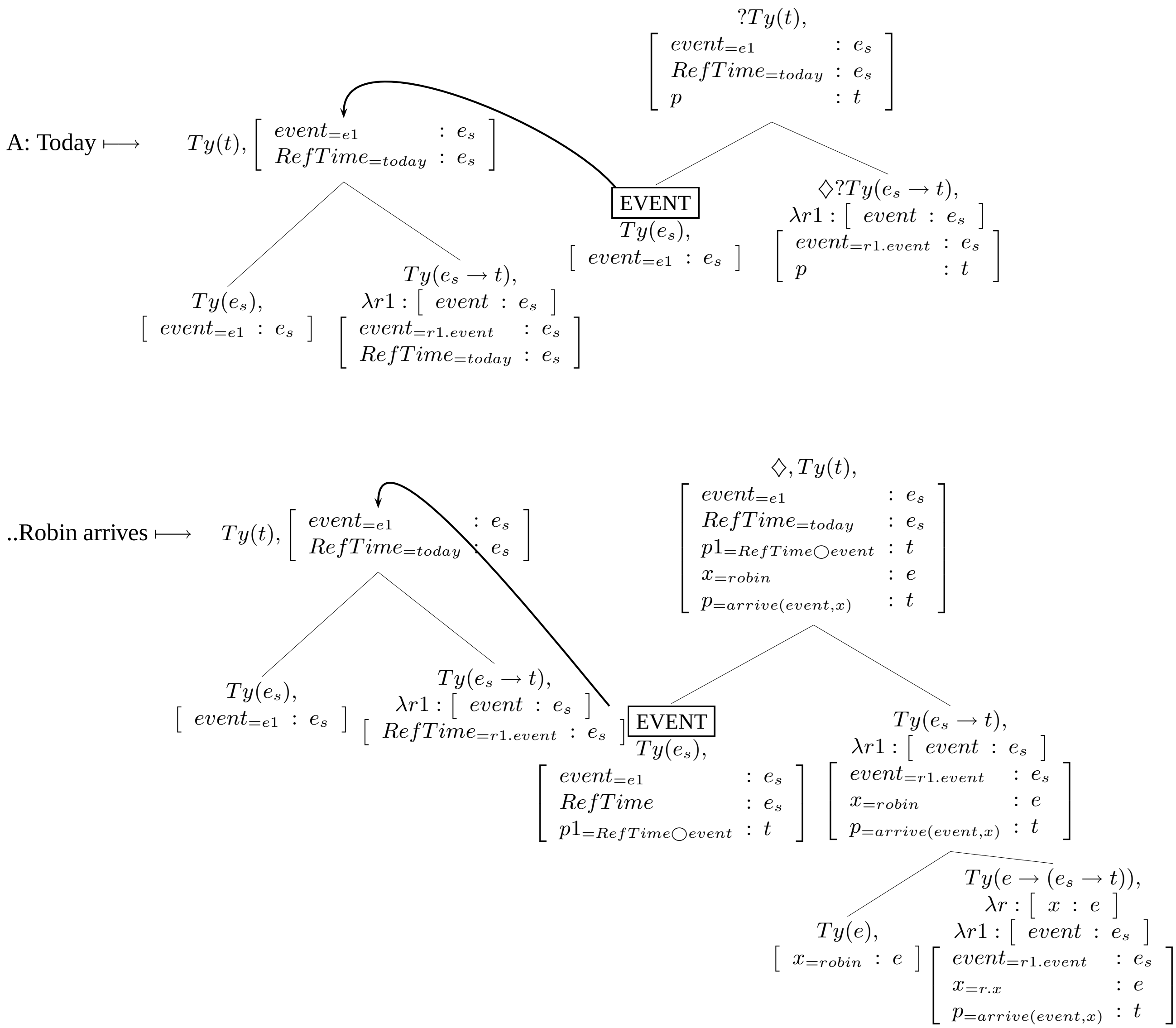

Figure 7: Processing “A: Today, Robin arrives”

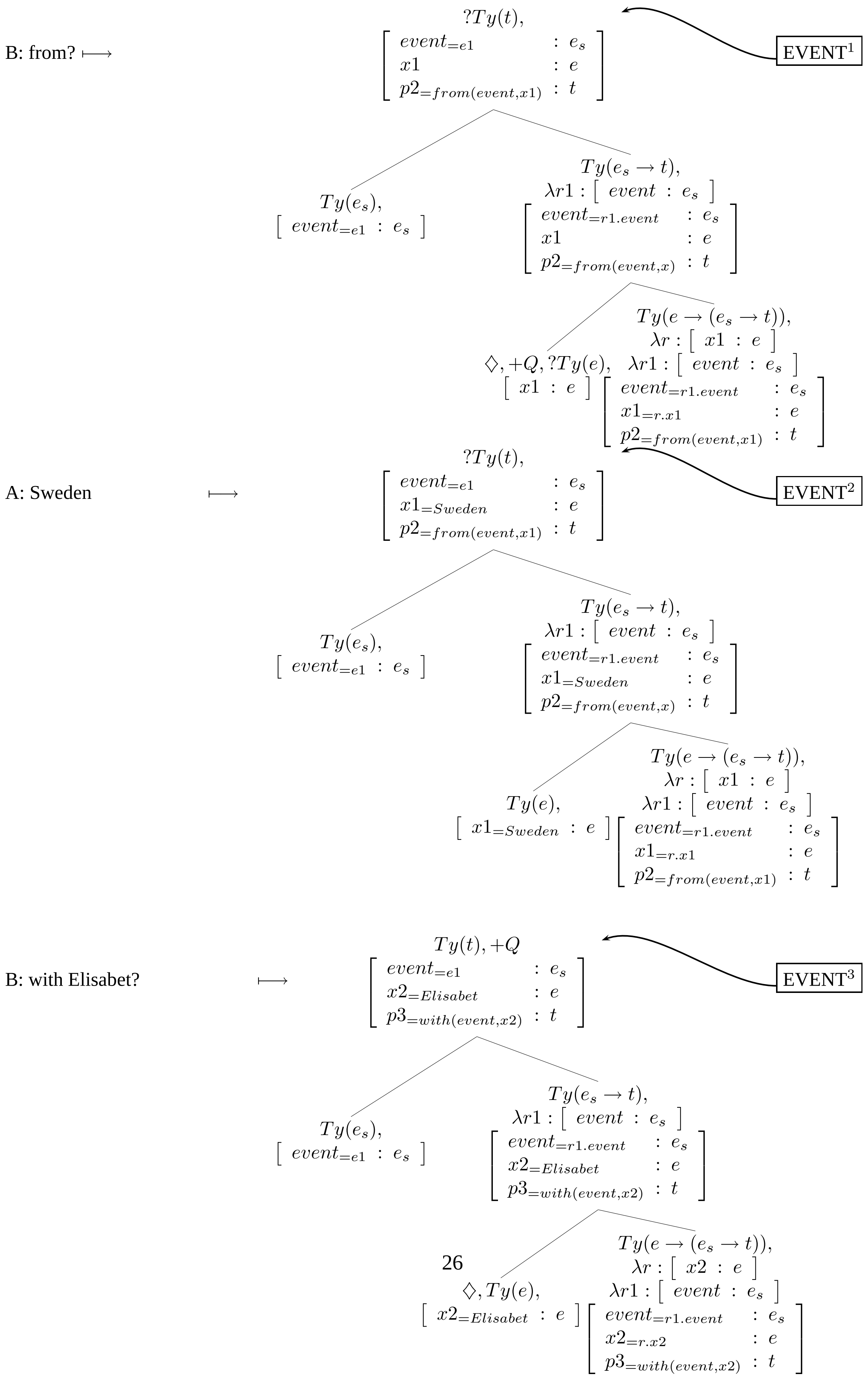

Figure 8: Processing Fragment (continued from Figure 7): “B: from A: Sweden B: with Elisabet?”

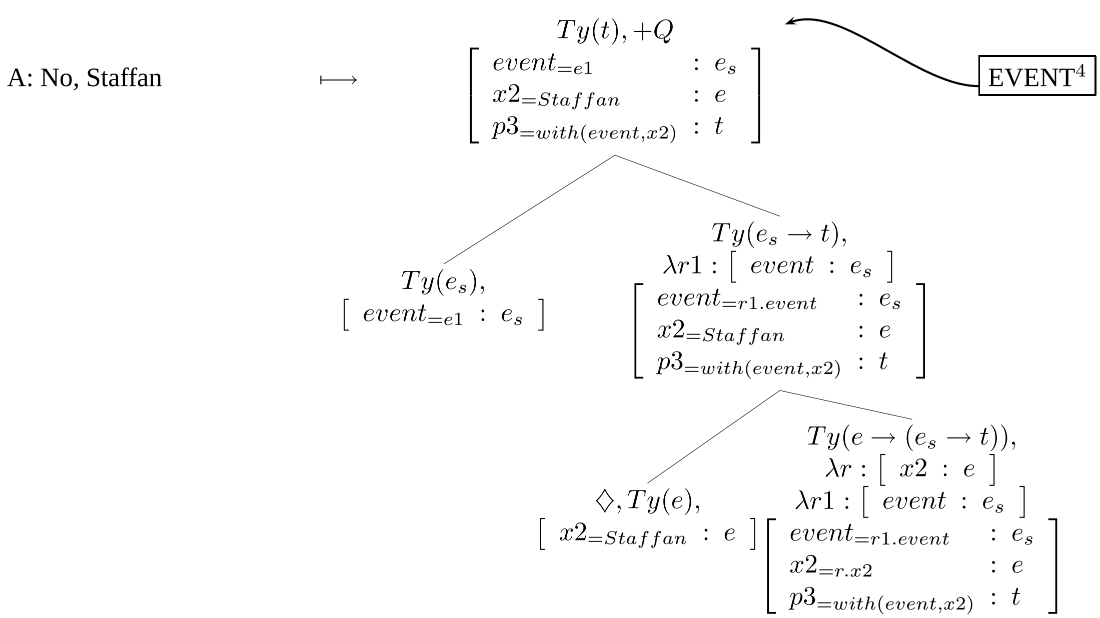

Figure 9: Result of processing “No, Staffan”: Other correction via backtracking along context DAG