3

On Coordination in Dialogue: Sub-sentential Talk and its Implications

ELENI GREGOROMICHELAKI, RONNIE CANN, AND RUTH KEMPSON

3.1. Introduction: brevity, communication, and meaning

According to a standard view of linguistic communication, a speaker who has a certain proposition (thought) in his/her mind manages to express it by forming and fulfilling the intention of getting the hearer to recognize that he/she has that thought. This transfer of propositions is mediated through the performance of speech acts (e.g., assertion) where the speaker selects a sentence in the expectation that, on the basis of this choice and the circumstances of utterance, the hearer will be able to infer the intended proposition. Grice's foundational work on speaker's meaning (a variety of so-called $\text{meaning}_{\text{NonNatural}}$) and principles for rationality and co-operation in conversation brought to the fore that there was a vast gap between what can be taken as provided by a grammar reflecting linguistic *competence* in the form of sentence-meanings paired with strings of the language, and the rich diversity of utterance interpretation, often only very indirectly related to what might be identified as some base linguistic meaning.

Over the subsequent years, the gulf between 'sentence meaning' and the concept of 'speaker meaning' has given rise to two primary paths of pragmatic research. Linguists either (as they saw it) broadly followed Grice in terms of defining the utterance content which a speaker intended to convey via normative maxims of co-operative behaviour (possibly via apparent violations of such maxims) or they pursued cognitive approaches (see, e.g., Sperber and Wilson 1986). And a few linguists advocated some mixture of these two (see, e.g., Levinson 2000). What is common to all such characterizations is their individualistic inferential nature. It is simply presumed

that utterance interpretation involves recognition of some (set of) propositions intended by the speaker.

This chapter shows the problems that elliptical phenomena, most prevalent in conversational dialogue, raise for these approaches, and argues for a distinct *language-as-action* perspective that penetrates deeper than usually assumed, that is, inside the mechanisms of the grammar itself. We assume as our point of departure a characterization of *ellipsis* as a form of brevity of expression (see Goldstein, Introduction to this volume), rather than as involving something being left out of an utterance that has to be reconstructed. Under standard views, ellipsis might be seen as required either through the conversational maxim of perspicuousness or via the various computations to minimize effort employed by the cognitive approaches. Yet, by retaining the view that utterance understanding invariably involves recovering a proposition, such pragmatic views require that the respective avoidance of prolixity or articulatory effort has to be compensated for by either encoded syntactic/semantic principles or application of inferential procedures, or possibly both, in order to yield the requisite 'complete' thought. In line with this, many varieties of ellipsis are postulated by researchers—syntactic, semantic, and pragmatic—each of these further subdivided. Embracing this diversity and incommensurability is then taken to justify the fragmentation of the knowledge involved in using language into (informationally encapsulated) modular components (see, e.g., Merchant (forthcoming); Martí 2009; Stainton 2006b). This view, which we attribute to a mistaken conception of the nature of NL grammars, we seek to challenge.

While nevertheless giving recognition to diverse ellipsis types within the grammar, Stainton (2006b) and Elugardo and Stainton (2001) have questioned the assumption that use of sentences is an essential precondition for successful communication (with ensuing debate with Merchant). To the contrary, they show that speakers can perform genuine speech acts from sub-sentential constituents, using inferential mechanisms, and without the need to first recover complete syntactic sentences or sentence contents. In this chapter, we show that examination of the extensive use of elliptical fragments in dialogue suggests a more radical view: the mechanisms of the grammar can themselves directly serve the same function of performing intelligible speech acts—which we characterize as *grammar-induced speech acts* (see Section 3.3)—without requiring either steps of inference or recovery of propositions. Since, as we shall argue, these mechanisms are low-level unconscious processes that do not plausibly involve propositional inference or standard notions of 'rationality' and 'co-operation', we shall conclude, more contentiously, that there is need to revise the assumption which others take to be the backdrop for pragmatic debates: that communication is achieved through manipulation of propositional intentions via (conscious or unconscious) conceptual reasoning processes. However, even more contentiously, we will argue that the data are also counterexamples to 'autonomous syntax' types of approaches (see, e.g., Merchant (forthcoming); Martí 2009). We shall

argue, to the contrary, that the mechanisms which constitute the core syntax are qualitatively continuous with those that ground the interpersonal practices that make manifest inferential effects that cannot be characterized strictly at an individualistic level of analysis (compare Brandom 1994; see Preston 1994). This conclusion will be reached by taking as our primary basis for modelling ellipsis resolution the effects of coordination between participants—a fundamental feature of conversational dialogue (see, e.g., Pickering and Garrod 2004). One such effect, brevity in forms of expression, results from mechanisms for context-dependent growth of interpretation which the processing system itself makes available without needing any inferentially—and hence propositionally—mediated co-operation by the dialogue partners. This will lead us to question the centrality of intention-recognition as the basis for utterance interpretation, and, as a corollary, to re-evaluate the presumed gulf between linguistic competence and language performance.

3.2. Ellipsis and the syntax–semantics–pragmatics interface

In current accounts, the consensus is that ellipsis is not a homogeneous phenomenon. Rather, it splits into syntactic types, semantically controlled bases for ellipsis, and pragmatic types, with only the latter type depending on context for its construal. The general background for both syntactic and semantic accounts is the methodology of conventional grammars which dictate the *sentence* as the unit of characterization: the only forms of ellipsis addressed here have been those where the ellipsis site can in some sense be reconstructed sententially:

(3.1) A: Have you seen Mary?
- (a) B: Mary? [(Are you asking:) Have I seen Mary?] [clarification]
- (b) B: No, I haven't. [No, I haven't seen Mary.] [VP-ellipsis]
- (c) B: Yes, and Tom too. [I have seen Mary and I have seen Tom.] [Stripping]

Thus (3.1a) can be understood as an echo of the original question, (3.1b) as the negative answer and so on. Indeed (3.1) illustrates a number of different ellipsis types. Each such type has been argued to be a separate syntactic phenomenon on the evidence of apparently different structural constraints governing their reconstruction as full sentential forms, operational at some level of abstraction at which certain structures and special null elements are present but are not pronounced (of which there is a long tradition, see, e.g., Hankamer and Sag 1976). Examples taken to illustrate this include *VP-ellipsis*, *stripping*, *gapping*, and *sluicing* (see Ginzburg and Sag 2000; Ginzburg and Cooper 2004 as representative of HPSG-style accounts; Merchant (forthcoming) as representative of minimalist (deletion) forms of account).

In other cases, not formally distinct from the above, a semantic explanation has nevertheless been argued to be required since, for a single antecedent form, ambiguity in interpretation arises:

(3.2) John checked over *his* mistakes, and so did Bill/Bill did too.

a. '*Bill* checked *Bill*'s mistakes' [sloppy]
b. 'Bill checked John's mistakes' [strict]

In such analyses, this is taken to reflect a process of abstraction over some content provided by the antecedent ('John checked over John's mistakes') creating distinct abstracts to apply to the content of the fragment in the elliptical conjunct. ('λx. x checked x's mistakes' / 'λx. John checked x's mistakes') (see Dalrymple et al. 1991, the point of departure for semantic analyses since). Hence, the semantic explanation involves a special-purpose semantic rule for resolving the content of such ellipsis sites. Under this conception, information from context then needs to be invoked to disambiguate the multiplicity of predicates generated by the semantic rule.

Beyond the syntax/semantics controversy, there are yet further cases where, it is claimed, there is no linguistically determined basis for assigning interpretation to the fragment. These are the cases observed by Elugardo and Stainton (2001) and Stainton (2006b), who argue that such cases have to be seen as speech acts performed without recourse to a linguistically determined sentential/propositional structure, it being non-linguistic aspects of context that determine the fragment's construal:

(3.3) Sanjay and Silvia are loading up a van. Silvia is looking for a missing table leg. Sanjay says:

'On the stoop'

According to Elugardo and Stainton, in such cases, first, it is evident that speakers mean more than what is articulated, that is, a word/phrase of semantic type <e> or <e,t> etc. In fact, the speaker conveys a proposition. Second, what is conveyed is not merely conversationally implicated, since it is hard to cancel the intended proposition, and because there are clear commitments to a set of truth-conditions induced by the sub-sentential form uttered. Thus, in such uses, we have a proposition asserted, despite a phonologically/syntactically/semantically 'incomplete' form.

Elugardo's and Stainton's insights aside, division of ellipsis into syntactic/semantic/pragmatic types by no means exhausts the range of competencies involved in successful manipulation of utterance fragments. Syntacticians/semanticists have until recently neglected dialogue phenomena, for example the phenomenon of clarification in Example (3.1), though this omission is gradually being rectified, with Ginzburg and colleagues leading this recent movement (e.g., Ginzburg 2012). Extending the Dalrymple et al. pattern, Ginzburg and Sag (2000), Ginzburg and Cooper (2004), Fernandez (2006), and Purver (2006) define multiple types of abstraction mechanisms to reflect distinct types of ellipsis, unifying characterizations over semantic, syntactic, and morphological specifications. But, in retaining the sentence-based methodology and a separate level of syntactic analysis, this approach is problematic. With such fragments having to be categorized as full sentences syntactically and

semantically, various predetermined speech act specifications are taken to be encoded (so, e.g., Example (3.1) will have to yield a proposition like 'Are you asking me if I've seen Mary (of all people)'/ 'Who are you referring to by *Mary*', etc). This has the immediate consequence of multiplying lexical/phrasal ambiguities for each expression serving as a fragment since each grammatical constituent will have to be assigned various distinct encoded interpretations and syntactic structures. Pragmatic principles or some formal regimentation of contextual effects are then invoked to resolve these as a disambiguation step. However, this strategy, which seeks to couple syntactic and semantic/pragmatic effects in a linguistically determined approach to the phenomena, cannot apply with full generality. Firstly, all the encoded speech act effects proposed to be hardwired in the grammar by Ginzburg and colleagues can be achieved without prior linguistic antecedents, a fact which undermines the claim that the grammar requires the multiplicity of levels assumed:

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

(3.5) B to A who is pointing at Harry: No, his sister [correction]

On the other hand, it is certainly the case that the use of fragments during interaction follows linguistically determined constraints which indicate their appropriate integration in some structured representation. This is more evident in languages with rich morphology. For example, although Elugardo and Stainton (2001) have shown that speakers can use fragments like the following in Example (3.6) to perform speech acts that do not presuppose the recovery of a full sentence, languages like German and Greek require that the fragment bears appropriate case specifications, otherwise it is perceived as ungrammatical:

(3.6) 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$_{NOM}$'

In such morphologically rich languages, fragments serving various dialogue functions also have to bear the appropriate case, gender, or other morphological features, while nevertheless having no linguistic antecedents:

(3.7) A is contemplating the space under the mirror while re-arranging the furniture and B brings her a chair:
A: tin karekla tis mamas?/*i karekla tis mamas? Ise treli? [clarification]
the chairACC of mum's/*the chairNOM of mum's. Are you crazy? [Greek]

In our view, such data do not indicate that the presence of grammatical constraints justifies a distinct level of representation for linguistically determined structure

(compare Merchant (forthcoming); Martí 2009; Stainton 2006b). Rather, such grammatical restrictions indicate that the grammar itself has to be defined appropriately in order to model constraints on interaction (see also Ginzburg 2012) at a unified representation that integrates input from various modalities.

One crucial attribute of this psycholinguistics-inspired syntactic modelling is the commitment to reflecting the incremental nature of processing, again illustrated most strikingly by dialogue. In conversation, utterances are often collaboratively constructed, with what is said by individual contributors being highly elliptical as they rely on the context in which the conversation takes place for their interpretation:

(3.8) 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
B: Just . . . talking about al you know alternative
D: On erm . . . renewable yeah
B: energy really I think . . .
A: Yeah (from BNC[1])

Given orthodox assumptions, it is far from obvious how to address this *split utterance* phenomenon. Standardly, the output of the grammar is a set of structures inhabited by complete sentences, as input to some performance theory for further enrichment. Upon such a view, none of these fragments will be included in the set of well-formed expressions, so a syntactic explanation has no obvious starting point. There are problems for semantic accounts also, for interruptions are possible at any point, and in some cases so early that no intended propositional content is as yet fixable. As a result, accounts following the Dalrymple et al. (1991) ellipsis resolution algorithm are not applicable without further assumptions:

(3.9) A. They X-rayed me, and took a urine sample, took a blood sample. Er, the doctor
B: Chorlton? [clarification]
A: Chorlton, mhm, he examined me . . . (from BNC)

To undermine further the uniform applicability of the semantic account, such subsentential switches involve speaker/hearer exchange of roles at any point, and across all syntactic dependencies (Purver et al. 2009), which indicates that a purely semantic account is not adequate:

[1] BNC refers to data found in the British National Corpus, see Purver et al. 2009.

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

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

(3.12) Gardener: I shall need the mattock.
Home-owner: *The* . . .
Gardener: *mattock*. For breaking up clods of earth. (from BNC)

(3.13) 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
(from Jones and Beach 1995, cited in Lerner 2004)

(3.14) A: or we could just haul: a:ll the skis in [the:] dorms
B: [we could] [haul all the skis into the dorm]
C: [hh uh hhuhhuh]
B: *which*
A: might work
B: might be the best (from Lerner 2004)

Within the syntax-based type of approach, Barton (2006) and Barton and Progovac (2005) claim that the Minimalist Program, unlike other frameworks, allows analyses in which Stainton types of sub-sentences can be licensed as independent constructions without sentential reconstruction. However, the split-utterance data, such as (3.8)–(3.14) above, go beyond what standard minimalist grammars generate as traditional constituents. In addition, as it has frequently been noted, there are also fragments that appear to be the result of transformations, hence the involvement of syntax cannot simply be to license context-free constituents:

(3.15) I simply hold up the letter, saying nothing, and you ask:
Where from? [Wh-movement]
(from Ludlow 2005)

3.3. Intention-based accounts of dialogue ellipsis

The challenge of modelling the full word-by-word incrementality required in dialogue has recently been taken up by Poesio and Rieser (2010) (*P&R* henceforth), who seek to explain the phenomenon through high-level accounts of coordination in

dialogue, namely, reasoning regarding intention-recognition. *P&R* set out a dialogue model for German, defining a thorough, fine-grained account of dialogue interactivity. Their primary aim is to model *collaborative completions*, as in (3.10). Crucially, their data comes from co-operative task-oriented dialogues where take-over by the hearer relies on the remainder of the utterance taken to be understood or inferable from mutual knowledge/common ground. Their account is an ambitious one in that it aims at modelling the generation and realization of 'joint intentions' which accounts for the production and comprehension of co-operative completions. The *P&R* model hinges on two main points: the assumption of recognition of interlocutors' intentions according to shared joint plans (Bratman 1992), and the use of incremental grammatical processing based on Lexicalized Tree Adjoining Grammar (LTAG).

The *P&R* account marks a significant advance in the analysis of such phenomena in many respects and, significantly, in that an incremental model of LTAG is assumed. Nevertheless, this account still relies on the assumption of a string-based level of analysis, for it is this which provides the top-down, predictive element allowing the incremental integration of such continuations. However, exactly this assumption would seem to impede a more general analysis, since there are cases where split utterances cannot be seen as an extension by the second contributor of the proffered string of words/sentence:

(3.16) Eleni: Is this *yours* or
Yo: *Yours*. [natural data]

In (3.16), the string of words that the completion yields is not at all what either participant takes themselves to have constructed, collaboratively or otherwise. Similarly, in (3.10) earlier, even though the grammar is responsible for the dependency that licenses the reflexive anaphor *myself*, the explanation for A's continuation in the third turn of (3.10) cannot be string-based as then *myself* would not be locally bound (its antecedent is *you*). Moreover, in LTAG (*P&R*'s syntactic framework) parsing relies on the presence of a head that provides the skeleton of the structure. Yet, as (3.9) and (3.14) indicate, utterance take-over can take place without the appearance of the head that determines argument dependencies (see also Purver et al. 2009; Howes et al. 2011). The data show that take-over can occur even across strict syntactic dependencies, e.g. in an antecedent-anaphor relation such as (3.10), a quantifier and its dependent variable as in (3.11), and in (3.17) between a Negative Polarity Item and its triggering environment, the question:

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

Given that such dependencies are defined grammar-internally, the grammar is needed to license such shared constructions. But string-based grammars cannot

account straightforwardly for many types of split utterances except by treating each part as elliptical sentences requiring reconstruction of the missing elements with case-specific adjustments to guarantee grammaticality/interpretability (as is needed in (3.16)–(3.17)). Given that such splits can occur at any point, a syntactic account would either necessitate processes of deletion of such power as to threaten theoretical viability, or the multiplication of types of syntactic analyses, hence indefinite structural homonymy, or both. Moreover, the rhetorical significance of one participant's taking-over the structure initiated by the other (co-construction) gets lost in such accounts (see later discussion of (3.13), (3.21), (3.26), and (3.27)).

Besides the problems engendered due to the assumption of an independent syntax, further considerations threaten the explanatory generality of *P&R*-style accounts. Even though the *P&R* model employs an incremental syntactic component, the account relies on the generation and recognition of the speaker's propositional intentions as the basis for the processing model. Yet firstly, in free conversation, as we saw earlier in (3.9), such fragments can occur before the informative intention—which is standardly defined as requiring a propositional object—has been made manifest. Secondly, unlike what happens in *P&R*'s task-oriented dialogues, many fragments do not involve straightforward participant co-operation or inference as to the speaker's intended utterance, hence a Gricean hierarchy of propositional intentions is not applicable (see Gregoromichelaki et al. 2011). For example, in (3.14) and the following examples, there is no reason to suppose that the continuation necessarily ensues only after B has considered A's intended utterance. B, who is in the process of parsing A's syntactic construction, just takes it over and appends material serving their own purposes:

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

(3.19) (A mother, B son)
A: This afternoon first you'll do your homework, then wash the dishes and then
B: you'll give me $20?

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

Moreover, such fragments can play multiple roles at the same time without any basis for characterizing one of them as 'indirect' or 'secondary' (e.g., the fragments in (3.10) and (3.16) can be simultaneously taken as question/clarification/completion/acknowledgment/answer). As argued in the Conversational Analysis literature, if full determination is even required, it will be effected by the recipient's uptake. However,

within the *P&R*-style accounts, such multi-functionality/indetermination would not be capturable except as a case of type-ambiguity or by positing some non-monotonic strategy that is able to apply even within the processing of an individual utterance. But such solutions, imposed by the need to employ explicit propositional intentions/plans on the part of the participants, are not desirable. Interlocutors do not have to be seen as motivated by antecedently present intentions as regards their speech act content. In fact, in some contexts, the vagueness of the speech act performed can be part of the pattern that sustains the interaction. For example, in psychotherapy sessions, invited completions have been argued to exploit the indeterminacy/covertness of the speech act involved to avoid overt/intrusive elicitation of information (Ferrara 1992):

(3.21) Ralph (therapist): Your sponsor before . . .
Lana (client): was a woman
Ralph: Yeah.
Lana: But I only called her every three months.
Ralph: And your so your sobriety now, in AA ::[(is)]
Lana: [is] at a year

As already illustrated above, contrary to the assumption that performance of speech acts must be mediated by intentions/plans embedding full propositional contents under pre-specified illocutionary force indicators, it seems that perfectly intelligible moves in dialogue can be achieved through what we characterize as *grammar-induced speech acts*. These are dialogue moves that are achieved simply by establishing *syntactic* conditional relevances,[2] that is, initiating a grammatical dependency which the interlocutor is invited to fulfil, thus providing for the performance of two or more (possibly incompatible) speech acts within one sentence. For example, completions might be explicitly invited by the speaker thus forming a question–answer pair:

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

(3.23) A: And they ignored the conspirators who *were* . . .
B: Geoff Hoon and Patricia Hewitt
(from Radio 4, Today programme, 6 January 2010)

(3.24) Jim: The Holy Spirit is one who <pause> gives us?
Unknown: Strength.
Jim: Strength. Yes, indeed. <pause> The Holy Spirit is one who gives us? <pause>
Unknown: Comfort. (from BNC)

[2] We borrow the term *conditional relevance* from the Conversation Analysis literature where it is used to indicate the type of sequential coherence that links the parts of *adjacency pairs*, see, e.g., Schegloff (2007).

(3.25) 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 . . . (from BNC)

As Lerner (2004) has shown, there are established practices for requesting elaboration exploiting the establishment of syntactic dependencies:

(3.26) Jack[3]: I just returned
Kathy: *from* . . .
Jack: Finland (from Lerner 2004)

Kathy's continuation has here a clear interpretation as a request for elaboration, further pinpointing the specific type of elaboration required. Unlike the standard completions studied by *P&R*, this type of response seems to reopen the turn of a prior speaker after it has apparently been completed by employing the initiation of a syntactic dependency. The performance of a further speech act (request for elaboration) is then achieved by employing the syntactic mechanisms that generate expectations as to what is to follow up (we call this feature of the processing mechanisms *predictivity*, see Section 3.5). The following exchange is characteristic of this dynamic: here, a doctor asks, *How ya doin' today*, to which the patient responds by a conventional 'small talk' phrase:

(3.27) Doctor: How ya doin' today.
Patient: .hh ? A:h pretty good, (0.5)
Doctor: *BU:t*, . . .
(cited in Lerner 2004)

The doctor's response then treats the patient's utterance as incomplete by initiating a continuation, *Bu:t*, whose own incompleteness serves as a request for further information regarding the health of the patient, hence their reason for their visiting the doctor. Here the doctor's response not only prompts for further information by a backward and forward looking continuation but, in doing so, shifts the type of the previous speech act from one where a conventionalized response to a greeting is offered to one where relevant medical information has to be provided.

The significance of these data is that such exchanges show the active involvement of the hearer in shaping the content of the dialogue in a way that does not warrant the primacy of the speaker's intention for the recovery of the significance of the speech act.[4] However, some such pre-specified 'joint' intention/plan is what drives the *P&R*

[3] In this case, Jack has been asked if he has just returned from a trip by someone who knew he had been away.

[4] For a way of modelling such retrospective construction of speech act content see Purver et al. (2010).

account of completions and many more accounts of coordination in dialogue (see, e.g., Grosz and Sidner 1986), despite the fact that such fixed joint intentionality is decidedly non-normal in free conversation (see, e.g., Mills and Gregoromichelaki 2010). One route for *prima-facie*[5] exclusion/minimization of intention computation would be to 'grammaticalize' such phenomena by encoding pre-specified speech act indicators accompanied by full propositional/sentential recovery. But, given the near-universal commitment to a static performance-independent methodology, approaches such as Ginzburg and Cooper (2004), Merchant (forthcoming), and even Elugardo and Stainton (2001), and Stainton (2006b), would require the assignment of explicit propositional contents even to those acts where conversational participants exploit the grammatical mechanisms to jointly achieve dialogue moves (*grammar-induced speech acts* see, e.g., (3.21)–(3.27)). However, it seems to us that this is neither necessary nor conceptually desirable. Since it is obvious that the grammar is crucially implicated in the licensing of all such data (see especially (3.10)–(3.17)), it seems that an appropriate performance-compatible architecture is what is required to provide a unitary account for all. This would provide an array of low-level mechanisms which underpin practices, whether ad hoc or conventional, enabling conversational participants to coordinate efficiently at the sub-conceptual, sub-propositional level, without their needing to explicitly represent the outcome and goals of their actions.[6] Thus our claim is that an appropriately conceived, psycholinguistics-inspired, grammatical framework provides the necessary means for modelling coordination in dialogue without necessarily involving the establishment of pre-specified communicative intentions/plans.

3.4. Use-compatible grammars and dialogue coordination

We have argued that the view emerging from dialogue data is that an appropriately defined model should be able to provide the basis for direct modelling of dialogue coordination as an immediate consequence of the grammar architecture. One such grammar is the Dynamic Syntax (DS) framework (Kempson et al. 2001; Cann et al. 2005). Uncharacteristically for grammars, DS is an action-based model of which the core notion is goal-directed incremental information growth/linearization following the time-linear flow of parsing/generation. Utterance contents, represented as binary tree-structures of predicate-argument form, are built up relative to a context which evolves in parallel, keeping a record of extra-linguistic information,

[5] *Prima facie* because intention computation has to then be invoked to resolve the resulting ambiguities.

[6] Nevertheless, explicit representation is not excluded either, as in many of the Elugardo and Stainton cases and especially when things go wrong or when there is a need to rationalize participants' actions. For the means of achieving this, without burdening the primary mechanisms of coordination see, e.g., Piwek (2011).

the sequence of unfolding partial tree-structures, and the actions used to build them. The process of building up such representations in context is what is taken to constitute NL syntax: syntactic constraints are modelled as procedures that define how parts of representations of content can be incrementally introduced and updated.

The general process is taken to involve building as output a tree whose nodes reflect the content of some utterance—in the simple case of a sentence uttered in isolation, a complete propositional formula:

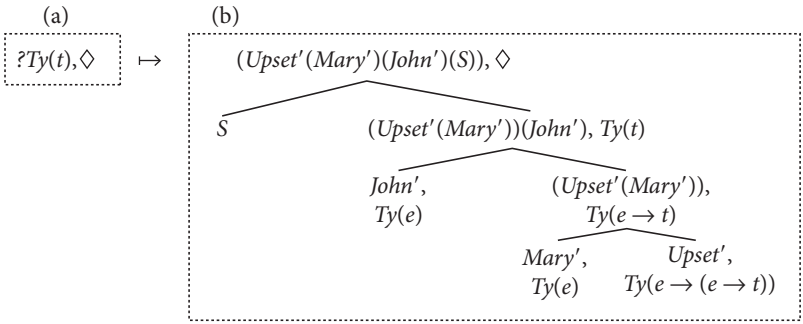

FIGURE 3.1. The outcome of parsing *John upset Mary*.

The input to this task, in such a simple case, is a tree that does nothing more than state at the root node the 'goal' of the interpretation process to be achieved, namely, to establish some propositional formula (?*Ty*(*t*) in Figure 3.1; goals are represented with *?* in front of annotations). For example, in the parse of the string *John upset Mary*, the output tree in Figure 3.1 to the right of the ↦ constitutes some final end result: it is a tree in which the propositional formula itself annotates the root node, and its various sub-terms appear on the dominated nodes rather like a proof tree in which all the nodes are labelled with a formula and a semantic type. These DS trees are invariably binary, and, by convention, the argument always appears on the left branch, and the functor on the right branch (a pointer, ◇, identifies the node under development). Each node in a complete tree is annotated not with words but contents, i.e. terms of a logical language (e.g., *Mary′*, *λx.Upset′x*), these being sub-terms of the resulting propositional representation (*Upset′(Mary′)(John′)* holds at index *S*). The parsing task is to use both lexical input, computational actions (e.g., *Introduction* and *Prediction* in step [1] in Figure 3.2) and information from context to progressively enrich the input tree satisfying all the sub-goals imposed (the satisfaction of goals introduced with *?* initially is indicated by removal of *?* when satisfied):

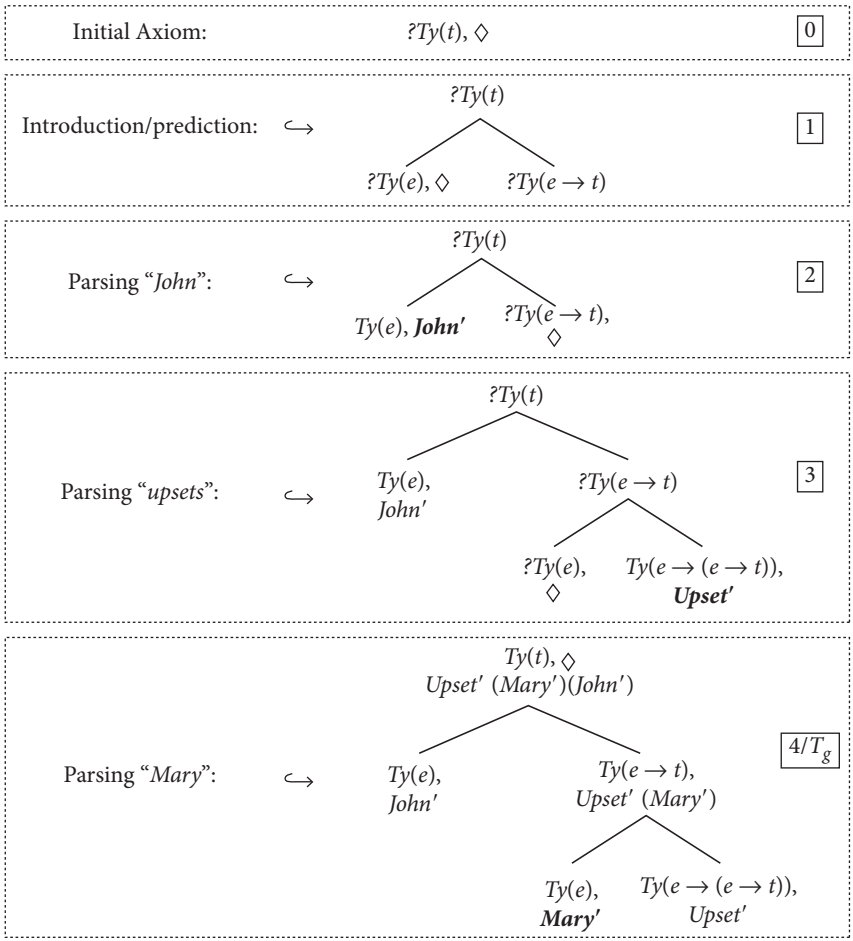

FIGURE 3.2. Incremental parsing/generation of *John upset Mary*.

These are the only representations constructed during processing, hence no distinct syntactic level of representation is assumed. Production follows exactly the same procedures, but with the added requirement of a subsumption relation to some richer 'goal' tree. For example, the tree T_g, the goal tree, shown in Figure 3.2 step 4, will be present from the beginning as the target of processing in case the speaker has planned a full proposition in advance. However, more partial trees can be assumed as targets in production, with the only requirement that the goal-tree is always at least one processing step ahead from the tree currently being processed.

As in DRT and related frameworks (see also Jaszczolt 2005), semantic, truth-conditional evaluation applies solely to these contextually enriched representations,

hence no semantic content is ever assigned to strings of words (sentences). The distinguishing feature of DS as compared to DRT is that this process of progressive building of semantically transparent structures is taken as core syntax: there is no other level of structure interfacing the phonological sequence and some ascribable content. Furthermore, all syntactic dependencies are seen in procedural terms including, in particular, the classical evidence for denying the direct correspondence between NL-structure and semantic content that led to accounts via transformations (see, e.g., Example (3.15)) (Kempson et al. 2001; Cann et al. 2005; Kempson et al. 2011a; and others following them). For example, 'movement' cases ('Who did you see'; 'The man who you saw') are analysed in terms of the initial projection of an underspecified dominance relation between the input provided by the WH-element and the predictively induced predicate-argument structure. Later update to a fixed dominance relation occurs at the point at which, in movement accounts, an associated 'empty category' is posited. Cases of so called 'syntactic/semantic ellipsis' may occur when the linguistic input includes anaphoric elements that have to be obligatorily enriched from the surrounding context. Anaphoric elements introduce *meta-variables*, symbolized as ***U***, ***V***, ***W***, along with goals triggering context search for their replacement with semantic terms. For example, a VP-ellipsis site will induce the introduction of a meta-variable in the place of the predicate which will have to be subsequently substituted from context:

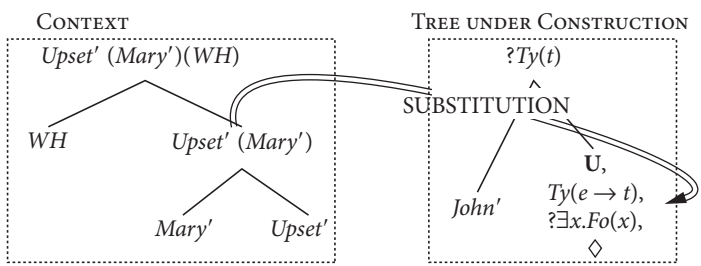

Figure 3.3. Q: Who upset Mary? Ans: John did. (strict readings)

The context may consist of structure induced by linguistic means relative to which the ellipsis site is linguistically determined to trigger the construction of a propositional content as in (3.1)–(3.2) and in processing the answer in Figure 3.3. But, given that context records both structure and already performed actions, reuse of either can yield distinct results: if representations of content are what is recovered from context, the result will be strict interpretations; if it is the actions that yielded such content that are retrieved from context, the result will be a sloppy interpretation, as those very same actions will then be re-applied at the fragment site to combine with whatever is available there (a new subject in the case of VP

ellipsis, see Kempson et al. 2011b for detailed analyses). Moreover, like pronominal anaphora, sometimes the contents that are being recovered will not have been provided by linguistic means, as in the Stainton cases, or even in cases of alleged 'syntactic ellipsis' where no linguistic antecedent is available (contra the prediction of Hankamer and Sag 1976):

(3.28) Uttered by a mother to her son as they stand facing the waves of a notoriously dangerous surfing beach:
'I *wouldn't* if I were you.'

Notwithstanding the emphasis on procedures, the structural properties of the emergent DS semantic trees are nonetheless crucial for accounts of dialogue phenomena. For example, it has been shown both by corpus research (Fox and Jasperson 1995) and experimental results (Eshghi et al. 2010) that repair processes in dialogue target primarily 'constituents'. Additionally, as we saw, the use of fragments during interaction follows syntactic constraints indicating their appropriate integration in some structured representation (see, e.g., (3.10)–(3.15) and (3.17)). As shown earlier in (3.6)–(3.7), this is more evident in languages with rich morphology. On the DS account, such morpho-syntactic particularities do not justify distinct levels of explanation, for the morphological information is defined as introducing constraints for appropriate integration in semantic tree representations: in particular, case information such as 'accusative' is taken to project a constraint that the content of an expression bearing this feature must occupy the appropriate argument position of a predicate. Hence such syntactic/morphological restrictions do not have to be taken as justifying a separate level of syntactic analysis assigning structure to strings of words (ie, sentences instead of contents). Modelling NLs as encoding constraints on growth of interpretation relative to context is exactly the assumption that allows the handling of dialogue phenomena such as split utterances in a straightforward manner, that is, as continuations involving genuinely jointly constructed contents. If, instead, a separate level of syntactic representation is insisted upon, such data can only be treated as fragments requiring propositional reconstruction or mechanisms overriding the morphosyntactic information they bear. This is because, as shown in Example (3.29) and earlier in (3.16)–(3.17), splicing together the two partial strings gives incorrect interpretations, since elements like indexicals have to switch form in order to be interpretable as intended, or for grammaticality:

(3.29) G: when you say it happens for a reason, it's like, it happened to get *you* off
D: off *my* ass
(from Clancy et al. 1996)

Instead of data such as those in (3.16)–(3.17) and (3.29) being indicative of language-particular levels of syntax/morphology, with ellipsis seen as ineliminably

heterogeneous, use of the licensing mechanisms both by a single speaker and seamlessly across interlocutors, as in a DS-style dynamic account, is what enables the unitary analysis of ellipsis. The two architectural features of DS that underlie this dynamicity and its direct licensing of partial sub-sentential constructs are *incrementality* and *predictivity*, features conventionally associated only with parsers (Sturt et al. 1996, 2005). Incrementality, that is, the licensing of sub-sentential elements as they become available in a time-linear manner, is an essential characteristic for the modelling of dialogue coordination. Firstly, the appropriate placing of items like inserts, repairs, hesitation markers, etc, has been shown to interact with the grammar at a sub-sentential level (Clark and Fox Tree 2002). Furthermore, dialogue phenomena like self-repair, interruptions, corrections (as in (3.5)), etc, rely on the incremental nature of both understanding and production in order to be modelled as making a timely contribution. But since, as we saw, the grammar must license such constructions, the elements it needs to manipulate must be partial/non-fully-sentential constructs. Because the syntactic licensing defined by DS is procedural and word-by-word incremental, fragments can be taken as just that, and not themselves sentential in nature. Accordingly, they may provide regular update to emerging partial structures irrespective of who has initiated these structures, as in the fragment interruptions in (3.8)–(3.9), or when the fragment is interpreted as an extension of a non-propositional structure given in context, as in (3.16)–(3.27).

Incremental integration of contents is coupled in DS with a general *predictivity/goal-directedness* in that the parser/generator is always predicting top–down structural goals to be achieved in the next steps (see Figure 3.2 earlier). Although, generally, the motivation for this type of architecture is efficiency considerations in parsing, in fact, coordination phenomena in dialogue can be seen to be exploiting this processing characteristic. For example, the turn-taking system (see, e.g., Sacks et al. 1974) seems to rely on the grammar, based on the predictability of (potential) turn endings. In this respect, recent experimental evidence has shown that this predictability is grounded on syntactic recognition rather than prosodic cues etc (De Ruiter et al. 2006); and further evidence shows that people seem to exploit such predictions to manage the timing of their contributions (Henetz and Clark 2011). More importantly for our concerns here, incremental planning in production, in combination with the predictivity of the parsing mechanism, allows the modelling of how the interlocutors interact sub-sententially to derive joint meanings, actions, and syntactic constructions taking in multi-modal aspects of communication and feedback. A DS-style predictive architecture for the grammar models these licensing mechanisms by means of the generation of goals to be achieved symmetrically by both the parser and the producer, the hearer/parser usually awaiting input from the speaker for fulfilling these goals. Such goals are also what activates the search of the lexicon ('lexical access') in generation in order

to recover a suitable NL word for the concept to be conveyed. As a result, an initial hearer/parser who achieves successful lexical retrieval before processing the anticipated linguistic input provided by the original speaker can spontaneously become the generator and take over. As seen in all cases (3.8)–(3.27) above, the original hearer is, indeed, using such a structural anticipation to take over and offer a completion that, even though licensed as a grammatical continuation of the initial fragment, might not necessarily be identical to the one the original speaker would have accessed had they been allowed to continue their utterance (as in (3.16)–(3.20)). And since the original speaker is licensed to operate with partial structures without having a fully formed intention/plan as to how it will develop (as the psycholinguistic models in any case suggest), they can integrate immediately such offerings without having to be modelled as necessarily revising their original intended message. By way of illustration, we take a simplified variant of (3.10) (for detailed analyses see Kempson et al. 2009; Purver et al. 2010, 2011; Gregoromichelaki et al. 2009, 2011):

(3.30) Ann: Did *you* burn
Bob: *myself*?

Here, the reconstruction of the string as *'Did you burn myself?' is unacceptable (at least with a reflexive reading of *myself*), illustrating the problem for purely syntactic accounts of split utterances. But under DS assumptions, with representations only of structured content, not of putative structure over strings of words, the switch of person is entirely straightforward. Consider the partial tree induced by parsing Ann's utterance 'Did you burn' which involves a substitution of the meta-variable (**U**) projected by *you* by the constant standing for the addressee/parser (*Bob′*):

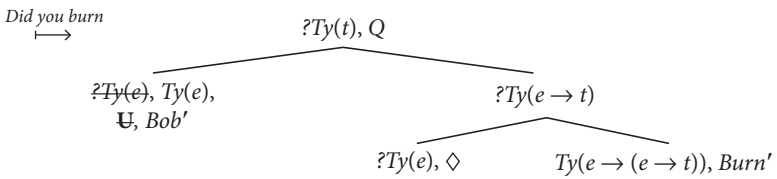

FIGURE 3.4. Ann utters: *Did you burn.*

At this point, Bob can complete the utterance with the reflexive, as what such an expression does, by lexical definition, is copy a formula from a local co-argument node onto the current node, just in case that formula satisfies the conditions set by the person, number, and also here the participant role of the uttered reflexive. So, in this case, the restriction is that the meta-variable stands for a local co-argument that is *currently* the speaker:

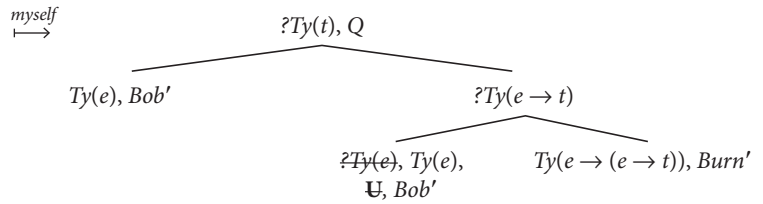

FIGURE 3.5. Bob utters: *myself*.

Hence, the absence of a 'syntactic' level of representation distinct from that of semantic representations allows the successful direct integration of such fragments through the grammatical mechanisms themselves, rather than necessitating their analysis as sentential ellipsis (for detailed analyses see Kempson et al. 2011b; Purver et al. 2010).

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 (Goodwin 1981), hence it is essential that the grammar does not have to license whole propositional units before semantic and pragmatic evaluation can take place. And this is the strategy DS adopts, operating directly with partial constructs whether induced by speaker or hearer: in either case, such constructs are fully licensed by the antecedently constructed context and thereupon integrated into the emergent semantic representation by updating it, without having to consider such fragments as sentences which happen to be elliptical, or as sentences which are in some sense not well-formed despite their success as utterances.

3.5. Concluding remarks: language as mechanisms for interaction

Given these results, in our view, the dichotomy between $language_S$ (language structure) and $language_U$ (language use) (Clark 1996) postulated in standard psycholinguistic models does not withstand the test of application in dialogue—the primary site of language use. Instead, linguistic knowledge has to be re-conceptualized as encompassing the update dynamics of communication, and has to be formulated in terms that are neither domain-specific nor encapsulated (contra Fodor 1983 and thereafter). It is not domain-specific because structural growth is defined over representations of content which integrate information derived from non-linguistic input. And it is not encapsulated because context-dependency phenomena, for example the construal of pronouns and ellipsis (see earlier Examples (3.1) and

(3.28) and Figure 3.3), are resolved online whether from contextual representations or from the construction process itself. Thus the grammar has been transformed into a specification of constraints all abstracted from the crystallization of action patterns derived from language use and wider cognitive or social considerations.

As a result, grammar-internal explanation of split-utterance data demonstrates that core dialogue activities can take place without any other-party meta-representation at all.[7] Hence, more generally, on this view, communication is not at base the intention-recognizing activity presumed by Gricean and post-Gricean accounts. Rather, speakers can be modelled as able to air 'incomplete' thoughts with no more than the vaguest of planning and commitments as to what they are going to say, expecting feedback to fully ground the significance of their utterance, to fully specify their intentions (e.g., Wittgenstein 2009: 337). Hearers, similarly, do not have to reconstruct the intentions of their interlocutor as a filter on how to interpret the provided signal; instead, they are expected to provide evidence of how *they* perceive the utterance in order to arrive at a joint interpretation. The sequential organization of conversation makes this a necessary feature: utterances are construed as simultaneously both forward- and backward-looking in terms of how they modify the context of interaction. This view of dialogue, though not uncontentious, is one that has been extensively argued for, under distinct assumptions, in the Conversational Analysis literature (e.g., Haugh 2012). According to the proposed DS model of this insight, the core mechanism is incremental context-dependent processing, implemented by a grammar architecture that reconstructs 'syntax' as a goal-directed activity, whose mechanisms are able to seamlessly integrate with the joint activities people engage in.

This then opens up a new perspective on the relation between linguistic ability and the use of language. Linguistic ability is grounded in the control of (sub-personal, low-level) mechanisms (e.g., Böckler et al. 2010) which enable the progressive construction of structured representations to pair with the overt signals of the language. The content of these representations is ascribed, negotiated, and accounted for in context, via the interaction among interlocutors and their environment. As a result, adopting but extending Stainton's (2006b) and Elugardo and Stainton's (2001) assumptions, the implications for semantics are indeed that independently specifiable contents and licensing have to be assigned to sub-sentential constituents. In addition, along with Elugardo and Stainton, we do not believe that dialogue activities (e.g., performance of speech acts) necessarily involve sentences or sentential contents. However, unlike Elugardo and Stainton, we do not think that the data justify the view that use of fragments necessarily involves mapping onto propositional contents either. Hence, we do not subscribe to the view that it is reasoning over speaker

[7] Though, of course, use of reasoning over mental states is not precluded either; such richer contexts and consequent derived implications are modeled via the construction of appropriately linked representations, whose mechanisms for construction are independently available in DS, see Purver et al. (2011).

intentions with respect to such contents that provides the essential basis for all utterance interpretation. Utterances can communicate content within the context in which they occur and give rise to successful interaction without the entertainment of any hypothesis about the interlocutor's mental state or intentions. From this perspective, constructing representations of the other participants' mental states, though a possible means of coordination, is by no means necessary.

Thus, from a general philosophical point of view, it seems to us that linguistic interaction is not essentially grounded in high-level inferential mechanisms or the discursive practices that underpin rationality ('sapience', contra Brandom 1994). Rather, it is low-level pre-conceptual mechanisms, envisaged by analysts as what they term the *grammar* of a language, which drive the coordination of participants in the construction process itself, hence prior to any possible justification on the part of the participating agents. This allows that the exchange of reasons remains as a significant component in an account of the epistemological significance of communication (see, e.g., McDowell 1980) but at a distinct level of description from the most basic account of coordinative activities. And here, a crucial insight of Brandom's—his 'expressivism' regarding the logical vocabulary—can be taken as a useful insight, though operating at a rather different level than he envisaged. Conversational participants also have the ability to 'make explicit' the practices afforded to them implicitly by the sub-conscious procedures when communication breaks down or, in general, when they need to verbalize/conceptualize the significance of their actions (for a similar account of practices at other higher levels of coordination see Piwek 2011). Some authors take these as evidence of the participants' explicit awareness of the speech acts, intentions, contents, practices, or conventions employed, and hence, they postulate such constructs as functionally relevant representations in modelling the achievement of coordination in dialogue (see, e.g., Stone 2004). However, in our view, this is the wrong tactic. It is the practices which are more basic; and explicit representations of any such high-level constructs result only from meta-reasoning or interpretive, reflective acts on the process of coordination (see also Mills and Gregoromichelaki 2010). It is on this approach to the linguistic system that, finally, we can explain the pervasiveness of the 'brevity' of utterances in dialogue—why there should be such extensive use of ellipses, interruptions, continuations, and other such discourse phenomena. If the basic mechanisms of direct human communicative interaction provide a means by which participants can jointly construct not just content, but also speech acts, and even intentions, then it is unsurprising that contributions to a dialogue will tend towards the minimum linguistically necessary to enable the interlocutors to do this using prior discourse, current parse states and external factors. Grice's brevity maxim thus appears to be not so much a rational demand internalized by participants, but the result of the low-level processing mechanisms that humans bring to bear in engaging in linguistic acts.