Cite this chapter as:
Gregoromichelaki E., Kempson R. (2016) Joint Utterances and the (Split-)Turn Taking Puzzle. In: Capone A., Mey J. (eds) Interdisciplinary Studies in Pragmatics, Culture and Society. Perspectives in Pragmatics, Philosophy & Psychology, vol 4. Springer, Cham

Joint Utterances and the (Split-) Turn Taking Puzzle

Eleni Gregoromichelaki and Ruth Kempson

Abstract This chapter argues that the occurrence of jointly constructed utterances (*split utterances*) in conversation has wide implications for current linguistic theories. Firstly, we show that standard formal syntactic and semantic/pragmatic theories are unable to cope with such conversational data due to the widely assumed competence/performance distinction. We then present some recent developments in the domain of formal models of dialogue, in order to assess whether they meet the design features that a general analysis of dialogue, and of the split-utterance phenomenon in particular, demand. We argue that what is crucial for such an account is incorporating both the physical and social situatedness of language use, combined with modelling the incrementality of linguistic processing, within the grammar formalism employed. In previous work, we have argued that the grammatical framework of Dynamic Syntax (DS) augmented with the flexible semantic representations made available by Type Theory with Records (TTR) meets these requirements. Accordingly, through the phenomenon of split utterances, we illustrate how the grammar itself needs to be seen as a holistic, action-based model of language use incorporating incremental interaction with context and flexible mechanisms of processing. These requirements are needed in order to deal, not only with what have been traditionally thought of as indexicals, but also with the representation of fine-grained sub-sentential utterance events, speech-act information, roles assigned to participants, etc. This stepwise interaction is necessary for a general account of how a speaker-change in mid-utterance affects the form and interpretation of linguistic elements. As a result, the incremental stance allows a natural characterisation of *split utterances* as continuations/interruptions, whereas, without it, the only recourse is the assumption of widespread ellipsis, mind-reading and multiple ambiguity of sub-sentential fragments. We then take a wider view of the data characterised as the Turn Taking Puzzle (Ginzburg 2012) by combining the phenomenon of split utterances with an account of the function of *why?* fragments ((Split-) Turn Taking Puzzle, STTP). On the basis of the STTP data, we argue that it is crucial for syntactic specifications and interpretation to interact with the modelling of the sub-sentential dynamics of the discourse-situation updates. From these interactions,

E. Gregoromichelaki (✉) · R. Kempson
King's College London, London, UK
e-mail: elenigregor@gmail.com

A. Capone, J. L. Mey (eds.), *Interdisciplinary Studies in Pragmatics, Culture and Society*, Perspectives in Pragmatics, Philosophy & Psychology 4,
DOI 10.1007/978-3-319-12616-6_28

we draw conclusions as to the significance of the STTP data for the design of grammar formalisms and dialogue models, as well as for the general conception of linguistic knowledge.

Keywords Dialogue · Speech acts · Discourse · Sub-sentential fragments

1 Introduction: Language Use in Interaction and the Shape of the Grammar

1.1 Standard Methodological Assumptions in Natural Language Modelling

A common position in the philosophy of language, largely adopted in the domain of formal semantics, has been the separation of the 'intentionality' of natural language (NL) and thought from:

a. the exercise of the capacities and epistemic resources that underpin perception and action
b. the environmental and sociocultural factors influencing and shaping these capacities.

On the basis of this separation, it has been assumed that an adequate theory of meaning will be given through a formal theory of 'truth' for NL sentences (see e.g. Davidson 1967; Larson and Segal 1995; Montague 1970). Such a theory for NL provides a system of finite resources that, internalised by an individual in a form of 'knowledge-that', enables the user of the theory to understand and produce every well-formed sentence of the language (*compositionality*). It is then claimed that, beyond these assumptions, if we turn to examine the employment of this abstract knowledge in realistic settings, i.e. in communication, we would inevitably be led to the conclusion that we

> 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, pp. 445–446, italics ours)

As a response to a similar worry in the domain of NL form (*syntax*), 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, p. 29, italics ours)

This claim is underpinned by philosophical/psycholinguistic views according to which the study of NL use does not provide for the isolation of a 'coherent system' of inquiry. Consequently, in formal and theoretical studies, various forms of abstraction are introduced to deliver a cleaned-up, idealised domain of data and theoretical tools for NL analysis. Modelling recursion in syntax and compositional-

ity in semantics, totally divorced from issues of processing and situated use, have become the sole criteria of adequacy for such theories and have motivated in turn a strictly individualistic/internalist methodology in the investigation of the nature of NL justified by a strict separation between the (modelling of) linguistic knowledge (*competence*) and the application of this knowledge in actual situations of language use (*performance*).

1.2 Challenges to Standard Methodologies

1.2.1 Situated Use and Semantics/Pragmatics

These methodological hypotheses have been called into question by several researchers interested in modelling the capacities underpinning NL use. In the domain of formal semantics, this has led to border disputes with pragmatics, in that it is no longer clear whether the separation between 'linguistically encoded' meanings and online, dynamically derived ones, in interaction with contextual factors, can be enforced. In this respect, there has long been work emphasising the role of linguistic underspecification in the process of deriving meaning in context (see, e.g. Sperber and Wilson 1995; Levinson 2000; Capone, this volume) and formulating notions of 'procedural meaning' that cannot be accommodated under truth-theoretic conceptions of semantics (e.g. Blakemore 1987). Further inadequacies of traditional truth-based theories have been highlighted by the 'dynamic turn' in semantics (Discourse Representation Theory (DRT, Kamp & Reyle 1993), Dynamic Predicate Logic (DPL, Groenendijk and Stokhof 1991) and related frameworks, see also Jaszczolt et al., this volume) that have drawn attention to the importance of conceiving meaning as updates to 'information states' rather than, statically, as sets of truth conditions (propositions) assigned to sentences. Nevertheless, all these approaches still concentrate on individual mental states modelled as autonomous representations that abstract away from the social and material circumstances of NL processing.

In contrast, recent efforts in formal semantics, inspired by work in Situation Semantics and DRT, have shifted attention away from a strict formulation of a truth theory as a theory of semantic competence to developing theories of semantic interpretation in context. For this purpose, a common representational system allowing the specification and seamless integration of multiple types of information has been sought. One recent articulation of this effort has been via the development of Type Theory with Records (TTR, see, e.g. Cooper 2005, 2012). TTR provides a transparent semantic representation format that can integrate both the low-level (sub-symbolic) perceptual information and underspecified, flexible meanings of NL expressions (see e.g. Larsson 2011). Such integration allows the modelling of how NL forms and meanings adapt to the discourse situation via the formalisation of an evolving, structured notion of the (multi-modal) context. Consequently, instead of adopting the assumption that the role of semantic theories is to assign truth conditions to decontextualised SENTENCES, in these approaches, attention has shifted to the modelling of situated UTTERANCES and speech acts. This has led to a significant expansion of the data deemed appropriate for inclusion in a formal theo-

ry 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). Another strand of this development, based on recent advances in developing compositional forms of DRT, is the PTT model (Poesio and Traum 1997, 1998; Poesio and Rieser 2010), which similarly expands the dynamic view of semantics to take into account underspecification of meaning resolved in context and language use in interaction. Similar developments have been taking place in the domain of syntax, to which we now turn.

1.2.2 Incrementality and Interaction in Syntax

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; Sturt and Lombardo 2005; Ginzburg and Cooper 2004; Kempson et al. 2001; Cann et al. 2005; Ginzburg 2012). In some of these models, instead of concentrating on the derivation of whole sentences as input to semantics, syntactic modelling involves the psycholinguistics-inspired commitment to reflect the incremental, situated and opportunistic nature of NL processing. This is because psycholinguistic experimental data, data obtained from natural conversations, as well as phenomenological intuitions, suggest that processing begins efficiently before the end of a sentence. For this reason, the effect of such incremental processing can be seen in real conversations where sentence-sized units are uncommon. However, under standard competence-performance assumptions, such speech data represent defective, degenerate NL uses which appear at best as either incredibly complex from a performance point of view or completely irrelevant from the standpoint of a competence theory. Indeed, conversational data do not display the idealised sentence-to-proposition format required by a competence grammar. Instead, they consist of 'fragments' (see e.g. turn 8 in (1) below) that are incrementally constructed and comprehended, and either then abandoned (turn 6, 7) or elaborated by the interlocutor (*split utterances*, see turns 3, 4, 5, 12, 14, 21):[1]

Example (1)

1. A: Instead of having <*name hidden*> <*unclear*> they had to come through the Dock Commission all of the men, they wanted so and so men for that boat, they used to come through to me.
2. B: Before that though, <*name hidden*> and <*name hidden*> [<*unclear*> had their own men]
3. A: [Had their own men]
4. B: unload the boats?
5. A: unload the boats, yes. They <*unclear*>
6. B: They were employed directly by

[1] The data that constitute the main focus of this chapter, here, split utterances, appear shaded in the examples.

7. A: That's right but they all came
8. B: <*name hidden*>?
9. A: They used to work say one week and have about a month off or go on the dole for a month.
10. B: So then what happened was, did the Dock Commission say you can't have your own men anymore?
11. A: That's right they had to go on a rota.
12. B: Run by the Dock Commission?
13. A: Run by the Dock Commission. See the dockers then all got together and they said right so many men for that job, so many for that job and that didn't matter who they were, they had to <*unclear*> their job, all the way round the dock.
14. B: Whether they wanted to go on that job or not?
15. A: Whether they want to go or not, they take their turn and the employer had to pay a percentage into the pool what those men earned, so when those men hadn't work at all they drew their money from the National Dock Labour Board.
16. B: Is this where the National Dock Labour Board came into existence?
17. A: That's how how they come into existence, yes <*name hidden*> he was a man what introduced that.
18. B: When was this?
19. A: Oh that's er, I would say about nineteen forty roughly <*clears throat*> Id say about nineteen forty that came in, might have been before that.
20. B: Before that then if they were ill
21. A: They get nothing.
22. B: Could they not get any welfare benefit?
23. A: No [BNC, H5H: 89–113]

In our view, split-utterance data demonstrate the radical context dependence of NL in conversation both on the syntactic and the interpretation side. Given orthodox assumptions, it is far from obvious how to address this context- and interlocutor-dependence. Standardly, the output of the grammar[2] is a set of structures defining well-formed complete sentences and propositional interpretations, which psycho-linguistic and pragmatic models then take as input to some performance theory for further enrichment. Upon such a view, none of the fragments above will be included in the set of well-formed expressions, so a syntactic explanation for their successful processing has no obvious starting point. They are problematic for semantic accounts also, since interruptions are possible at any point, and in some cases so early that no intended propositional content is as yet determinable. On the other hand, any account that overrides the grammar has to contend with the fact that such sub-sentential switches involve speaker/hearer exchange of roles across all

[2] We use here the notion of *grammar* common in formal semantics, where it consists of a syntactic component and a semantic component. We seek to redefine this notion of "grammar" to a more holistic model that includes pragmatics, and any relevant processing issues.

syntactic dependencies (Purver et al. 2009; Gregoromichelaki et al. 2011), which indicates that licensing has to occur via the same mechanisms that enable canonical intra-sentential licensing:

Example (2) Joe: We were having an automobile discussion….
Henry: discussing the psychological motives for
Mel: drag racing in the streets. [Sacks 1992, pp. 144–145]

Example (3) Helen: When I left you at the tube earlier, I went home and found my boyfriend…
James: In bed with another woman. Shit! [Sliding Doors]

Example (4) Louise: No a Soshe is someone who [is a carbon copy of their friend.
Roger: [drinks Pepsi.

Example (5) Ken: Instead my grandmother offering him a drink, of beer, she'll say ["Wouldju-"
Louise: ["Wanna glassa milk?"

In terms of pragmatic accounts, in traditional individualistic theories of speech acts (e.g. Searle 1969), speakers fulfil their communicative intentions by performing illocutionary acts embedding complete propositional contents. In this fulfilment, the interlocutor is modelled as a component of the speaker's knowledge, a factor shaping the content and form of the utterance only through the speaker's representation of what the SPEAKER perceives as being their mutual knowledge (*common ground*). However, as the data in (1)–(5) indicates and research in interactional linguistics (see e.g. Arundale 2008) and psycholinguistics has demonstrated, NL utterances and contents in real interactions involve incremental sub-sentential processing, situatedness (Mey, this volume) and feedback (Goodwin 1979) and, thus, in various senses co-construction by several interlocutors. Contra the single sentence/proposition methodology, utterances, like various other events in conversation (e.g. even silences), are always interpretable in the local sequential conversational environment which not only provides for their coherence in that particular sequence but also affects how the meaning of terms in the currently processed utterance is derived (see e.g. Schegloff 2007).

From these points of view, meaning is not inherent in individual-internal complete propositional thoughts delivered via speech acts performed turn-by-turn by interlocutors. As psycholinguistic studies have shown, the mechanisms that sustain interaction between individuals contribute in a crucial way to the development of meaningful exchanges. For example, Schober and Clark (1989) found that conversational partners who were given the means of interacting with a speaker had a different quality of understanding than overhearers who lacked this possibility, even though, from an external point of view, the information conveyed through linguistic means was exactly the same. In addition, language use in conversation is highly dependent, moment-to-moment during the interaction, on integrating and combining inputs from several senses comprising non-verbal behaviours and features of the physical environment (multi-modality). For example, in face-to-face communication, there is a tight linguistic and embodied synchronization between speakers and listeners (Pickering and Garrod 2004, 2012), with constant feedback

loops jointly determining the course of the utterance as it unfolds via verbal and non-verbal signals (Goodwin 1979, 1981, 1995). Conversational participants follow each other's utterances and behaviours incrementally, perceiving and acting in the discourse situation where elements acquire variable meanings according to their temporal appearance in the string of words (and not just under some overarching action to be completed at transition-relevant places, as claimed by Conversation Analysis accounts).

2 Language as Action, Plan-based Approaches

In order to cope with such data, a number of researchers have recently developed a wider conception of grammar as a component of a more holistic model of utterance interpretation and production. Embracing Austin's observation (1962) that NL use is a form of action, and, more specifically, joint action, as illustrated by the highlighted data in (1)–(5), NL understanding is subsumed under general models of action interpretation (e.g. Bratman 1990, 1992, 1993). The defining characteristic in such analyses of the significance of action is treating speaker intentions as *plans* and demonstrating how a speaker's utterances can be assigned structure and meaning in terms of the plans those utterances serve (see e.g. Grosz and Sidner 1986). Plans also link speech acts with non-linguistic behaviour and the environmental contextual constraints. Within such a research context, the challenge of modelling the full word-by-word incrementality required in dialogue has recently been taken up by Poesio and Rieser (2010).

Poesio and Rieser seek to explain the phenomenon of (a subcategory of) split utterances through adopting the assumptions of the planning model, namely, reasoning involving intention-recognition. They 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 (2)–(5). Crucially, their data come from co-operative task-oriented dialogues (e.g. video-recorded experiments where the participants are required to build something together).[3] In these cases, takeover by the hearer relies on the remainder of the utterance taken to be understood or inferable from a store of mutual knowledge (*common ground*). The Poesio and Rieser account aims at modelling the generation and realization of 'joint intentions', sharing of which is what, in their view, underlies the production and comprehension of collaborative completions.

Unlike standard formal semantic models which focus on a truth-based semantics for sentences and a view of common ground as a repository of mutually believed propositions to support inference (as in e.g. Stalnaker 1979), the first distinctive feature of Poesio and Rieser (2010) is the assumption—derived from ideas developed in Situation Semantics (Barwise and Perry 1983) and Clark (1996)—that the

[3] The significance of intention recognition even in task-oriented dialogue experiments has been disputed (see Mills and Gregoromichelaki 2010).

common ground representation also includes the *discourse situation*, i.e. the context of the conversation itself. Along with the mutually accepted truth-evaluable content of utterances, information about the discourse situation is recorded in a unified representation, a discourse representation structure (DRS), modelling each participant's 'information state' at each point in the dialogue. The occurrence of utterances of sub-sentential constituents is recorded in this representation as the occurrence of events in a certain temporal order (*micro-conversational events*) which thus become part of the common ground. The occurrence of these micro-conversational events leads to immediate updates of the participants' information states with the initiation of semantic and pragmatic interpretation processes (Larsson and Traum 2000; Stone 2004) following the specifications of the grammar. As regards pragmatic integration, in this model, speech acts are conceptualised as events too, termed *conversational events*, since just like any other events, they can serve as the antecedents of anaphoric expressions:

Example (6) A: You're an idiot.
B: *That* was uncalled for. [*that*: A insulting B]

More innovatively, speech acts are also viewed as components in a *joint plan* (Bratman 1992; Clark 1996) whose establishment and recognition drive the actions of speaker and hearer. Consequently, the Poesio and Rieser (2010) modelling of collaborative completions hinges on two main points: the assumption of the necessary recognition and adoption of the interlocutor's intentions according to the shared joint plan, and the use of incremental grammatical processing based on Lexicalized Tree Adjoining Grammar (LTAG). According to them, data like (2)–(5) require that the hearer, who knows the intention of the speaker and shares their utterance plan, produces a continuation that the speaker themselves would have provided otherwise.

Poesio and Rieser's thorough and detailed account of completions marks a significant advance in the analysis of such phenomena in many respects and, significantly, in that an incremental model of LTAG is adopted. As Gregoromichelaki et al. (2012), Eshghi et al. (2010, 2012) argue, this is a welcome approach since, in contrast to claims in Conversational Analysis research (e.g. Lerner 1991), the data show that takeover can occur anywhere in an emerging utterance, even across strict syntactic dependencies, e.g. earlier in (2) a preposition and its object, in (4) a relative pronoun and the rest of the relative clause, in (5) the verb and its propositional object, and below in (7) an antecedent-anaphor relation, and in (8) between a Negative Polarity Item and its triggering environment, the interrogative indicator:

Example (7) A: I heard a shout. Did you
B: Burn myself? No, luckily.
Example (8) A: Have you mended
B: any of your chairs? Not yet.

Given that such dependencies have to be defined grammar-internally, the grammar is unquestionably needed to license such shared constructions. Nevertheless, the Poesio and Rieser account cannot deal exactly with those crucial data. This is

because it still relies on the assumption of a string-based level of analysis, in that the grammar includes a distinct level of syntactic representation that provides a tree structure whose nodes are inhabited by words of the language. Sharing of utterance plans will generate identical string-syntactic representations for each interlocutor, and this allows the incremental generation and integration of other-initiated continuations. However, exactly this assumption threatens the generality of the 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*:

Example (9) Eleni: Is this yours or…
Yo: Yours. [natural data]

In (9), as well as in (7)–(8), the string of words (*sentence*) that the completion yields is not at all what either participant takes themselves to have constructed, collaboratively or otherwise. In (7), even though the grammar is responsible for the dependency that licenses the reflexive anaphor *myself*, the explanation for A's continuation cannot be string-based as then *myself* would not be locally bound (its antecedent is *you*). Moreover, in LTAG (Poesio and Rieser's syntactic framework), parsing relies on the presence of a *head* that provides the skeleton of the predicate-argument structure. Yet, as (1).3, (1).4, (1).12, (1).21 and (4), (7) indicate, utterance takeover can take place before the appearance of the head that determines argument dependencies (see also Purver et al. 2009; Howes et al. 2009, 2011). So, string-based grammars cannot account straightforwardly for many types of split utterances except by treating each part as an elliptical sentence requiring reconstruction of the missing elements with case-specific adjustments to guarantee grammaticality/interpretability (as is needed in (8)–(9)). Given that such splits can occur at any point, as we have shown, an ellipsis account would either necessitate processes of deletion and reconstruction of such power as to threaten theoretical viability (see, e.g., Morgan 1973), or the multiplication of types of syntactic analyses, hence indefinite structural homonymy (Stainton 2006), or both. Moreover, the rhetorical significance of one participant's taking over the structure initiated by the other (co-construction), instead of starting a new utterance, gets lost in such accounts (Gregoromichelaki et al. 2013b).

Besides the problems engendered due to the assumption of an independent string-based syntactic structure, further considerations threaten the explanatory generality of Poesio and Rieser-style, plan-based accounts. Their account relies on the generation and recognition of the speaker's propositional intentions as the basis for the explanation. Yet, in free conversation, such fragments can occur before the informative intention—which is standardly defined as requiring a propositional object—has been made manifest. Unlike what happens in Poesio and Rieser's task-oriented dialogues, many fragments do not involve straightforward participant co-operation or inference as to the speaker's intended utterance. For example, in the following, there is no reason to suppose that the continuation necessarily ensues only after the hearer has considered some propositional whole derived from the speaker's intended utterance (termed as *hostile completions* (13) or *devious suggestions* (14) in Gregoromichelaki et al. 2011):

Example (10) Helen: I, I'm sure you're not a nutcase or a psycho or anything, it's just that, um I'm not, I'm not that good at, um you know, um…
James: Constructing sentences? [Sliding Doors]

Example (11) Helen: I love this bridge. My great grandfather helped to build it. I often come and… stand on it when I want to, um…
James: Build a bridge? I'm sorry [Sliding Doors]

Example (12) Connie: Clarence, I am looking for you! Where are you? I want to talk to you! Clarence?
<Connie bangs hard on cupboard's door where Clarence is hiding>
Clarence: Ah, Connie, splendid! Erm… Heard you calling. Wasn't able to find you, so I thought, what a capital idea to…
Connie: Fling the servants' shoes around? [Blandings: Pig-hoo-o-o-o-ey! BBC2 14/1/13]

Example (13) (A and B arguing:)
A: In fact what this shows is
B: that you are an idiot

Example (14) (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?

Example (15) 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]

As Gregoromichelaki et al. (2011, 2012) argue, the hearer, who is in the process of parsing the speaker's syntactic construction, just takes it over and appends material serving their own purposes. The significance of these data is that such exchanges show overtly the active involvement of the hearer in shaping the content of the utterance, thus providing evidence that the primacy of speaker's intention for the recovery of the significance of the speech act is not a warranted theoretical assumption. However, some such pre-specified 'joint' intention/plan is what drives the Poesio and Rieser 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).

Further evidence against such plan/intention-based explanations comes from elliptical clarification questions which can query covert goals of an utterance. Such evidence is provided by Ginzburg (2012) to which we now turn.

3 The Interactive Stance on Grammar and *Why* Questions

Like Poesio and Rieser (2010), Ginzburg (2012) provides a holistic model that, while maintaining formal semantics' standard concerns, e.g. a notion of compositionality, seeks to integrate previously neglected aspects of utterance interpretation and production in a dialogue competence model. However, an important aspect of Ginzburg's model is that it takes the demands of language use in interaction as being built directly into the grammar to a much larger extent than does the Poesio and Rieser account. For example, in various analyses, he shows that, often, the conventional meaning of a word or syntactic construction involves reference to notions such as 'current issue under discussion', 'conversation initiation', 'acknowledgement of understanding' or 'ask intended reference of the interlocutor's utterance'. In order to provide analyses for such elements, like Poesio and Rieser, Ginzburg offers a model of context that assumes that the common ground in conversation not only includes a store of the interlocutors' common knowledge/beliefs but also all the facts related to the *discourse situation,* including facts about the form of utterances that have occurred, their grammatical types, phonology, syntax as well as semantics. As in Poesio and Rieser, illocutionary acts and utterance (*locutionary*) acts are treated in a uniform manner, i.e., as events whose occurrence is recorded in the common ground. However, unlike the reliance of Poesio and Rieser on shared plans, Ginzburg, instead, takes seriously the potential for misunderstanding, rejection and correction, which leads to detailed modelling of the divergences of each interlocutor's information state at various points in the dialogue. This divergence is directly built into the model of the context and has direct interaction with the grammatical specifications. In order to implement this, each interlocutor's information state is partitioned into a 'public' and a 'private' part. Each interlocutor's (version of) the public part, termed the *Dialogue Gameboard,* can be distinct at various points in the conversation according to whether the interlocutor assumes the role of either 'speaker' or 'hearer'. The private part includes beliefs not considered mutual and the plans and purposes underlying each conversational move (which are the factors driving the progress of the conversation in the Poesio and Rieser model).

Contrary to the Poesio and Rieser methodology, Ginzburg provides evidence from corpus data regarding clarification requests that plan recognition and joint acts are not a necessity for understanding an utterance and making it part of the common ground. The same data justify the need to differentiate information states (context) according to participant role in conversation. These data involve two uses of questions employing the utterance of fragmentary *Why?*-interrogatives. The first use, which Ginzburg calls *direct-why* interrogatives, is illustrated in (16). Here either the speakers themselves or the interlocutors can pose an elliptical *why*-interrogative to request an explanation for a fact introduced into the common ground by some previous utterance:

Example (16) A: Bo left yesterday.
A/B: Why? [: *Why did Bo leave yesterday?*]

The second use, termed as why_{meta}, is illustrated below in (17):

Example (17) Cherrilyn: You got a pound?
Jessica: Why? [: *'Why does Cherrilyn ask if Jessica has got a pound?'*]
Cherrilyn: ch I mean in change
Jessica: no. [BNC, KBL, cited in Ginzburg 2012]

The interpretation of such elliptical why_{meta} interrogatives involves a formalisation in the grammar of the phenomenon that Ginzburg characterises as 'metacommunicative' interaction, i.e. talk about the communicative process itself. A crucial element in this account is the adoption of TTR (see Sect. 28.1.2.1) as both the semantic representation language and the language in which the syntactic rules are formulated. This provides a uniform representational format allowing the reification of token locutionary and illocutionary speech-act events as they occur in the dialogue and the imposition of conditions on their occurrence (licensing). These representations of reified speech events can then be used as parts of the content of metacommunicative clarifications. For example, the event of Cherrilyn asking if Jessica has got a pound in (17) is recorded in the context (information state; as in cases like (6) earlier). It can then be used as the target argument of *why* ellipsis, i.e. asking why it has occurred, what was the interlocutor's plan that motivated its occurrence. Thus, posing and understanding why_{meta} queries involves interpreting the elliptical utterance *why?* as a query regarding the reason behind a speech act that has been performed recently, rather than querying the speech act's content as in *direct-why* constructions. So, according to Ginzburg, why_{meta} queries are used to clarify the goals underlying an utterance, i.e. the unpublicized intentions of the speaker, or in terms of the Poesio and Rieser (2010) account we discussed earlier, the plan motivating a speaker's utterance.

This analysis provides then the requisite argument against assuming that recognition of underlying intentions or plans in discourse underpins successful utterance understanding. First of all, in various corpus research, it has been shown that clarification requests regarding the recognition of goals (intentions) are only a minute percentage in comparison to other types of clarification, e.g. those regarding difficulty with intended reference or confirmation that a word has been heard successfully. This undermines the Poesio and Rieser account or any other accounts based on the planning model (e.g. Grosz and Sidner 1986, where plan structure underpins discourse structure). This is because, if the execution of a joint plan was the force driving communication, such clarifications would be crucial and they would be expected to occur as frequently as all other cases. It could be argued that such intention recognition does not pose any problems; after all, the participants are

engaged in a joint plan, according to Poesio and Rieser. However, a second piece of evidence Ginzburg provides undermines this. All other types of clarification, or repair in general, tend to be almost invariably local to the problematic utterance, i.e. occurring in the next turn (except where nested repairs have to be performed in order, i.e. as a 'stack'). Accordingly, Ginzburg treats the factors targeted by such local clarifications as pertaining to necessary contextual enrichments for an utterance to be comprehended (to be 'grounded'). In contrast, for why_{meta} clarifications, Ginzburg shows various attested cases where the successful integration of an utterance is accomplished, but where the *why?* query about intentions arises later in the conversation, after the problematic utterance has been comprehended and appropriately responded to:

Example (18) Norrine: When is the barbecue, the twentieth? (pause) Something of June
Chris: Thirtieth.
Norrine: A Sunday.
Chris: Sunday.
Norrine: Mm.
Chris: Why? [= '*Why do you ask when the barbecue is'*]
Norrine: Becau Because I forgot (pause) That was the day I was thinking of having a proper lunch party but I won't do it if you're going out. [BNC, KBK, cited in Ginzburg 2012]

Example (19) Cherrilyn: Are you still (pause) erm (pause) going to Bristol (pause) on Monday?
Fiona: Dunno.
Cherrilyn: No?
Fiona: I dunno. Doubt it, why? [= '*Why do you ask if I'm going to Bristol?'*]
Cherrilyn: I just wondered.
Fiona: Why?
Cherrilyn: I just wondered! [BNC, KBL, cited in Ginzburg 2012]

In the why_{meta} clarifications above, the to-be-clarified utterance has been adequately comprehended and responded to; therefore, the target of the clarification cannot be an essential parameter in integrating utterances to the common ground.

In terms of grammatical analysis, given the otherwise standard syntactic model adopted (a variant of Head-driven Phrase Structure Grammar, HPSG), different types of *why*-interrogatives are treated as hard-wired, distinct *constructions,* instead of linguistically underspecified elements, which is one of the crucial benefits which a TTR implementation, in our view, enables. Thus, Ginzburg postulates special grammatical types for fragmental *why*-interrogatives: in a rather unorthodox

fashion, an adverbial *why?* occurrence maps to a verbal phrase so that it can be assigned a sentential meaning, i.e. a meaning involving a proposition-like object.

The semantics that Ginzburg assigns to *direct-why* constructions, as in (20) below, involves interaction with recent facts recorded in the context. The meaning of a *why?* clause targets a fact *f* in the common ground and results in the question: ? λr.*Cause(r, f)*,[4] i.e. the speaker of such an elliptical *why*-interrogative seeks to clarify what the cause *r* is for some *f* in the common ground. But not any fact will be an appropriate antecedent. Ginzburg argues that the felicitous interpretation of such elliptical *why*-interrogatives requires that the set of facts (FACTS)[5] in the common ground is differentially structured according to each fact's relative saliency so that such facts can serve as antecedents to propositional anaphora. This is because of the interpretation of *why?* in data such as the following:

Example (20) A: Terrible weather recently.
B: But it's nice and sunny today.
A: Yes. Why?
[= '*Why is it nice and sunny today*?' But not: *'Why have we had terrible weather recently?'*]

The interpretation of *why?* above in (20), as targeting the most local fact that has just been introduced in the common ground, shows, according to Ginzburg, that facts in the common ground are structured dynamically according to their saliency at each point in the conversation. This is handled by the postulation of a category TOPICAL that only includes facts that have just been accepted or queries that are currently under discussion in the conversation. The facts stored in TOPICAL change as the conversation proceeds and Ginzburg provides a modelling of such dynamics.

The analysis of such structures has further consequences from an interactional point of view. A standard assumption in formal semantics is that context (*common ground*) is viewed as an abstract entity to which both participants, speaker and hearer, have common access. However, examining the so-called elliptical why_{meta} questions, Ginzburg draws different conclusions. Beyond the partitioning of each participant's information state to 'public' (Dialogue Gameboard) and 'private' parts, he argues that there has to be also a distinction according to what semantic objects are salient for the production of an elliptical utterance according to whether a participant is either speaker or hearer. This is motivated by the phenomenon termed as the Turn-Taking Puzzle (TTP, Ginzburg 1997, 2012), which is illustrated below:[6]

[4] Despite the fact that Ginzburg uses the predicate *cause,* he talks about explanation as regards the content of the query which, in at least most cases, will involve, in our view, a notion of 'reason' rather than 'cause'. Note also that *question* is the semantic object contributed by, among others, interrogatives and employed in acts of *querying*. Questions are analysed uniformly as λ-abstracts in Ginzburg's account.

[5] Although it is crucial for Ginzburg's model to distinguish various semantic objects like propositions, facts, questions etc., for simplicity, consistency and brevity of presentation here we avoid to make these distinctions as they do not affect the general argumentation.

[6] The symbol # in front of an utterance/interpretation indicates pragmatic unacceptability.

Example (21) A: Which members of our team own a parakeet?
B: Why?
[(a) # *'Why own a parakeet?'* (b) *'Why are you asking which members of our team own a parakeet?'*]

Example (22) A: Which members of our team own a parakeet? Why?
[(a) *'Why own a parakeet?'* (b) # *'Why am I asking this*?']

According to Ginzburg, the reading in which *why*? queries the reason behind the just performed speech act (the why_{meta} reading, (21)(b) and (22)(b)) is only available when the turn changes and the new speaker B uses the *why*-interrogative (as in (21)). This reading is not available when the original speaker A keeps the turn (as in (22)). However, Ginzburg argues, this is not simply due to coherence or plausibility as regards the posing of such why_{meta}-interrogatives of oneself, as this reading is available when expressed by non-elliptical means:

Example (23) A: Which members of our team own a parakeet? Why am I asking this question?

Example (24) A: Are you in cahoots with Tony? Why am I asking this?

Rather, this is because, according to Ginzburg, contexts are structured differentially for speaker and hearer, in that they do not have equal access to the salient semantic objects available in the context for ellipsis resolution. Moreover, this context dependence poses much higher demands in the structure of the context record than well-known cases of indexicality (*I*, *you*, *now*, *here*) since what is crucial here is also who made the PREVIOUS utterance.

However, despite the fact that Ginzburg claims that the restrictions affecting *why*? ellipsis resolution concern 'semantic objects', he does not reflect this solely in the semantics. Instead, he enshrines it in the syntactic component as a syntactic ambiguity, with two distinct structures for each interpretation of such fragments: *bare-why-clause*, for *direct-why* as in (16), (20) and (22), and why_{meta} clause for the ones in (17)–(19) and (21). The derivation of both such clauses makes crucial reference to the context, i.e. the contents of the structured common ground. However, the licensing of why_{meta} in particular ensures that the agent of the illocutionary act queried is distinct from the current user of the *why*-interrogative.

Notwithstanding these differences, both *why*-constructions involve unorthodox rules that map the adverbial *why* directly to a verbal phrase, in order to allow such fragments to head a sentence and stand as independent structures, as imposed by the need to mesh with other grammar rules. This is because Ginzburg assumes, as do Poesio and Rieser (2010), that the grammar is based on hierarchical syntactic structuring that concerns the string level, i.e. sentences. It is then because of this assumed level of NL structure, despite Ginzburg's claim that such constructions are *not* syntactically elliptical, that the grammar must, nonetheless, involve unmotivated category-changing structural rules in order to successfully derive an appropriate semantics for such fragments. As a consequence, despite the substantial enrichment of the grammar to ensure integration of illocutionary force for each utterance event, interaction with context, and characterisation of metacommunica-

tive interaction potentials, the syntax still remains a component of the model qualitatively distinct from semantics. Various semantic/pragmatic-syntactic mismatches then require postulating structural ambiguities in order to make each component internally consistent. As a result, in our view, the potential that the employment of TTR representations affords, that of defining and resolving underspecified linguistic elements by combining them with elements from the context, is overlooked in favour of constructional ambiguity.

4 Grammar as Mechanisms for Incremental Interaction: Dynamic Syntax-TTR

In contrast, a more radical alternative concerning the status of the syntax/semantics components of the grammar is proposed by Dynamic Syntax (DS, Kempson et al. 2001; Cann et al. 2005). DS is a psycholinguistically inspired 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 and representations. This model eschews a string-syntactic level of explanation and implements the assumption that grammatical constraints are all defined procedurally in terms of the progressive development of representations of content ('information states'), with partial interpretations emerging step-by-step during social interaction on a more or less word-by-word basis. In the view we sketch here, this is a variant which combines DS with the TTR framework (Cooper 2005, 2012; DS-TTR), which captures directly the fine-grained dynamics of dialogue, as well as the potential for underspecification and enrichment (Purver et al. 2010). In the next section, we set out the case for abandoning standard conceptions of the grammar (i.e. syntax-semantics mappings). Ultimately, we argue, this involves a reconceptualisation of what NL knowledge consists in, namely irreducible 'knowledge-how', rather than propositional 'knowledge-that'. We then present the application of the resources of this model to an account of split utterances and, finally, to the combination of split utterances with Ginzburg's TTP (illustrated earlier in (21)–(23)) in order to explicate how the constructional view does not generalise so that relevant data remain unaccounted for. On this basis, we then draw conclusions as to the appropriate format of a fine-grained integrational model of NL use, which, in our view, incorporates various aspects of the interface with perception, action and sociality in a single architecture.

4.1 Linguistic Knowledge: The View from the DS-TTR Perspective

Standardly, the formulation of grammars abstracts away from 'performance', i.e. processing and pragmatics, as it is assumed that use of NL presupposes the ontologically and conceptually prior specification of propositional knowledge regarding

a syntactic theory and a theory of meaning. Thus, *syntax* is confined to the licensing of sentence strings as a means of delimiting the set of well-formed sentences of the language. As we saw earlier, for Ginzburg (2012) such an assumption motivates the necessary assignment of a sentential category to adverbial fragments like bare *why*- interrogatives, while in Poesio and Rieser (2010) this is what prevents their account from treating data like (7)–(9) as genuine co-constructions of a single sentence. As regards *semantics*, standardly, interpretation is defined as the application to the set of structured strings of a Tarski-inspired truth theory yielding propositions as denotations, this being the interface point at which the contribution of the grammar stops and pragmatics takes over. Despite their many innovations, linguists like Ginzburg (2012) and Poesio and Rieser (2010) seek to preserve these basic aspects of this conception of semantics, by distinguishing the constraints of syntax, semantics, and context, despite the unified representations assumed. From a philosophical point of view, neo-Davidsonians (e.g. Larson and Segal 1995) further assume that knowledge of NL includes tacit propositional knowledge of this truth theory; this tacit knowledge is what enables individuals to produce and interpret speech appropriately in interaction with others possessing the same tacit knowledge.

Consequently, instead of modelling the mechanisms enabling the joint actions individuals engage in during interaction, such theories concentrate in delivering DESCRIPTIONS of such actions, expressed as various propositional speech-act characterisations. As a consequence of this stance, classical truth-based semantic theories have enshrined Frege's context principle (Frege 1884) 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. In our view, this is what motivates the necessity in most dialogue models, e.g. Ginzburg (2012) and Poesio and Rieser (2010), to analyse fragments of various semantic types that occur in dialogue as mapping to proposition-like semantic objects that correspond to explicit paraphrases of the perceived effects of such fragments. Standard speech-act theories have also embraced this view (see e.g. Searle 1969, p. 25). One of the reasons behind this stance is that the basic units of NL understanding are taken to be speech acts with propositional contents, 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 (Gregoromichelaki 2013b). For the same reason, even pragmatic models like Relevance Theory (Sperber and Wilson 1995) can only deal with propositions as providing sources of 'relevance', hence fragment analyses that employ this type of approach necessarily resort to propositional expansions again (e.g. Stainton 2006).

However, Davidson himself acknowledges that the individualistic psychological basis of this explanation of NL knowledge 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, p. 16)

Under the standard competence-performance assumptions though, it is unclear how orthodox syntactic/semantic models can deal with the modelling of meaning as deriving from 'interacting groups' since, because of their fragmentary, non-sentential nature, dialogue data, like the ones illustrated in (1)–(24) earlier, are delegated as secondary, recalcitrant and degenerate uses to performance. Hence, a puzzle ensues: on the one hand, various researchers are now admitting that meaning originates in interaction; on the other, real interactions appear to furnish data that are incompatible with the postulates of standard theories of NL structure and meaning.

In our view, the problem standard syntactic theories have in dealing with dialogue data can be traced to the assumption that it is sentential strings and propositional readings that constitute the output of the grammar (compatible with the philosophical/semantic views that adopt Frege's context principle), along with the attendant methodological principle debarring any attribute of performance within the grammar-internal characterisation to be provided. According to the DS perspective we take here, the problem starts with the overall requirement placed on NL models to provide accounts of 'communication', which is a concept still carrying the implications of the 'code model', according to which propositional messages are constructed in the mind of one interlocutor and then transmitted and decoded by the other. And this is an assumption that permeates most current accounts in formal semantics/pragmatics. Instead, within DS, we propose to reformulate the remit of grammars as the modelling of a set of unencapsulated, subpersonal mechanisms for action 'coordination', i.e. the meshing of (linguistic and non-linguistic) actions to achieve efficient joint performance, without necessarily requiring that explicit propositional representations have to be derived for the conceptualisation of speech acts or other actions as each step of coordination is being achieved. Crucial for such a conception of a processing model of the dynamics of coordination is that knowledge-how is involved at all levels of analysis, including 'syntax'. Consequently, the standard view, the bifurcation of syntax and semantics/pragmatics, is rejected by DS. Instead, via employing a procedural architecture modelling joint action during language use, DS conceives of 'syntax', hence meaning, as underpinned by two features usually associated solely with psycholinguistic models of parsing/production, namely (a) incrementality and (b) fine-grained radical context dependence. These two 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. 2005; papers in Kempson et al. 2011a, b; 2012a, b; Gregoromichelaki 2006; Kempson et al. 2012; Gregoromichelaki 2013a, b). DS is formulated as a system which crucially involves:

- an action-based architecture that models dynamically the development of unitary representations integrating multiple sources of contextual information,
- word-by-word incrementality and predictivity within the grammar formalism, and
- speaker/hearer mirroring and complementarity of processing actions.

We will not go into the details of the formalism and the computations here;[7] for our purposes it suffices to look more closely at how this perspective, when applied to dialogue modelling, sheds new light on dialogue puzzles: the phenomenon of split utterances seen earlier in (1)–(5) and (6)–(15), which we take up in Sect. 4.2, and the interpretation of *why?* fragments (earlier in (16)–(24)), as we will see in Sect. 4.4.

4.2 Incrementality and Predictivity in the Grammar Induce Split Utterances

Instead of deriving sentence structures and propositional meanings, the DS grammar models the word-by-word processing of NL structures in context. For language use in conversation, this is a crucial explanatory factor since many of its features rely on such incremental production and comprehension. For example, the frequent occurrence of clarification requests in conversation (Ginzburg 2012 inter alia) shows that utterances can be processed and understood partially without having to map a sentential structure to a full proposition (contra Ginzburg 2012). Moreover, it has been shown that in conversation, the positioning of items like inserts, repairs, hesitation markers etc. is not arbitrary but systematically interacts with grammatical categories at a sub-sentential level (see e.g. Clark and Fox Tree 2002 inter alia). In addition, hearers display their comprehension and assessments of the speaker's contribution sub-sententially as the utterance unfolds, through *back-channel* contributions like *yeah*, *mhm*, etc. (Allen et al. 2001). And speakers shape and modify their utterance according to the verbal and non-verbal responses they receive from hearers as their turn unfolds (Goodwin 1981). Hence, the grammar must be equipped to deal with those in a timely and integrated manner, i.e. by providing syntactic licensing and semantic interpretation online. In addition, the turn-taking system (see, e.g. Sacks et al. 1974) seems to rely on the grammar, as it is based on the predictability of (potential) turn endings in order for the next speaker to time appropriately their (potential) entrance; in this respect, experimental evidence has shown that this predictability is grounded mostly on syntactic recognition rather than prosodic cues etc. (De Ruiter et al. 2006). More importantly for our concerns here, we have argued that, since the grammar manipulated by both interlocutors in dialogue is a set of reactive and anticipatory actions, the role of the hearer is not passive but, instead, actively responsive and complementary to the speaker's actions (Gregoromichelaki et al. 2011). Thus, in contrast to intentional planning models like Poesio and Rieser (2010), seen earlier in Sect. 2, we argue that incremental production induced by subpersonal grammatical mechanisms is adequate to account for how the interlocutors interact sub-sententially in dialogue to derive joint actions, meanings and syntactic constructions, taking in multi-modal aspects of the environment and feedback, a fact claimed to be a basic characteristic of interaction (Goodwin 1981).

[7] We cite throughout the publications where the relevant formal details can be found.

The DS model assumes a tight interlinking of NL perception/production in that the grammar simply consists of a set of licensed actions that both speakers and hearers have to perform in synchrony in order to interpret or produce a step-by-step mapping from phonological strings to semantic representations consisting of formulae in the lambda-calculus.[8] As in DRT and related frameworks (see also Jaszczolt 2005; Jaszczolt et al., this volume), semantic, truth-conditional evaluation applies solely to these contextually-enriched representations, hence no semantic content is ever assigned to structures inhabited by elements of 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'. Unlike standard syntactic models, in DS, there is no intermediate level of syntactic structuring. Hence, strings of words are not assigned hierarchically organised constituency as phrases or sentences. Such constituency is considered in DS as epiphenomenal on the function-argument semantic relations as typified in the lambda-calculus analyses of NL meanings. In consequence, all syntactic dependencies have been reformulated 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 (long-distance dependencies, binding, quantification, etc.). Such phenomena have been shown to be explained by incorporating underspecification and its resolution within the syntactic dynamics (see e.g. Kempson et al. 2001; Kempson et al. 2012a, b; Cann et al. 2005; Gregoromichelaki 2006, 2011, 2013a).

According to DS, both speaker and hearer perform the processing steps incrementally, but, perhaps, in diverse contextual environments since the cognitive circumstances of each agent might be distinct. Given the fine-grained incremental DS architecture, efficiency (as well as psycholinguistic evidence) dictates that processing is not only strictly bottom-up, guided solely by the NL string, but also driven by *predictions* ('goals'). These are expectations imposed by either the procedures associated with NL elements ('lexical actions') or system-generated as general top-down computational goals to be achieved in the next steps. Simplifying for presentation purposes, for example, in English, with its characteristic subject–verb–object (SVO) structure, a general computational goal will ensure that parsing/production starts with the expectation of a subject first, followed by a predicate afterwards. The lexical entries for transitive verbs will introduce not only the conceptual content associated with the word but also the prediction/expectation that an argument, the object, will follow immediately afterwards. And likewise for all other regularities occurring in English or any other NL 'syntactic' structuring: the actions associated with words will induce the appropriate processing steps and predictions. Thus, parsing in DS incorporates elements of production through the generation of predictions for what will ensue next. On the other hand, production exploits the parsing mechanism in that licensing of the generation of each word relies in checking that the string so far produced can deliver a conceptual representation that

[8] The language of the epsilon calculus is combined with the lambda calculus in order to deal with quantification, see Kempson et al. (2001); Gregoromichelaki (2006, 2011).

accords with the (partial) conceptual structure the speaker attempts to verbalise. As a result, speaker and hearer roles involve mirroring of each other's actions (Gregoromichelaki and Kempson 2013; Pickering and Garrod 2012).

As speakers and listeners simulate the actions of each other, the fulfilment of syntactic/semantic goals (predictions) is essential at each incremental step, subsententially, for both parser (hearer) and generator (speaker) and can be satisfied by either, whether on the basis of the other interlocutor's input or by recourse to the processor's own resources and context. As no structure is ever assumed to be derived for the sentence string, no whole-string 'grammaticality' considerations ever arise. Hence, fragments that can be processed by fitting into a structure that is already in the context are licensed directly, NOT as elliptical, without the assumption that they need to be enriched to a propositional type:

Example (25) A: Who left?
B: John
C: with Mary, yesterday.

Split utterances are then unproblematically processable and, in fact, a natural consequence of such a fine-grained bidirectional incremental system: As goals are constantly generated by the grammar, to be achieved symmetrically by both the parser and the producer, the hearer/parser can await for input from the speaker in order to fulfil these goals. However, according to the grammar, such goals are also what activates the search of the lexicon ('lexical access') in production in order to recover a suitable NL word for the concept to be conveyed. As a result, an initial hearer/parser who achieves a successful lexical retrieval before processing the anticipated NL input provided by the original speaker can spontaneously become the producer and take over verbalising the continuation of the utterance instead.

For this reason, from an interpretational point of view, DS predicts a much wider range of split-utterance types than Poesio and Rieser (2010) with their standard syntax-semantics articulation. The Poesio and Rieser model is perhaps able to cope with the type of split utterances termed *collaborative completions* as in (1)–(5) earlier. However, it is very much less compatible with the many other types of continuations in conversation. As (10)–(15), repeated below, show, such completions by no means need to be what the original speaker had in mind, so an account of their generation does not need to involve prediction at the message or semantic levels, just the ability of the original hearer to go on from the point at which parsing has stopped:

Example (10) Helen: I, I'm sure you're not a nutcase or a psycho or anything, it's just that, um I'm not, I'm not that good at, um you know, um…
James: Constructing sentences? [Sliding Doors]

Example (11) Helen: I love this bridge. My great grandfather helped to build it. I often come and… stand on it when I want to, um…
James: Build a bridge? I'm sorry [Sliding Doors]

Example (12) Connie: Clarence, I am looking for you! Where are you? I want to talk to you! Clarence?
<Connie bangs hard on cupboard's door where Clarence is hiding>
Clarence: Ah, Connie, splendid! Erm… Heard you calling. Wasn't able to find you, so I thought, what a capital idea to…
Connie: Fling the servants' shoes around? [Blandings: Pig-hoo-o-o-o-ey! BBC2 14/1/13]

Example (13) (A and B arguing:)
A: In fact what this shows is
B: that you are an idiot

Example (14) (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?

Example (15) 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]

From this point of view, coordination in dialogue does not require replicating thoughts in the interlocutors' minds but, instead, enabling each other to go on with the activity they are engaged in. Thus, these cases (termed *hostile continuations* or *devious suggestions* in Gregoromichelaki et al. 2011; Gregoromichelaki and Kempson 2013) and many others go against Levinson's (2012) assumption that mindreading is necessarily involved in NL action-coordination: there is no reason to suggest here that, before interrupting, the listener first figured out the speaker's plan, then derived the expected continuation, then rejected it, then figured out a new plan which resulted in an alternative continuation which he/she then produced, while the original speaker went through the reverse process in order to comprehend and integrate this continuation. Such data then cast doubt on the pervasive Gricean assumption, a residue of the code model, that in all successful acts of communication, the speaker must have in mind some definitive propositional content which they intend to convey to their hearer, whose task, conversely, is to succeed in grasping that particular content.

But even in cases where the continuation appears to be a 'guessing' of the original speaker's intention, in fact, as (7)–(9), repeated below, show, the string of words (sentence) that the completion yields is not at all what either participant would have planned from the beginning, so these cannot be licensed by a standard grammar:

Example (7) A: Have you mended
B: any of your chairs? Not yet.

Example (8) A: I heard a shout. Did you
B: Burn myself? No, luckily.

Example (9) Eleni: Is this yours or…
Yo: Yours. [natural data]

To generalise over all cases, (1)–(5) and (7)–(15) earlier, we have argued that the original hearer is simply using 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. And since the original speaker is licensed to operate with partial structures without necessarily having a fully formed intention/plan as to how the utterance 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 (for detailed analyses, see Eshghi et al. 2010, 2011; Gargett et al. 2008, 2009; Gregoromichelaki et al. 2009, 2013a; Kempson et al. 2011a; Purver et al. 2006, 2009, 2011).

4.3 Speech Acts in DS-TTR

Unlike standard assumptions in Poesio and Rieser (2010) and other planning models, where an intended speech act, assertion, query, request, etc. has to be recorded in the common ground to achieve the appropriate understanding of an utterance, we believe that such necessary derivation is not part of the usual interpretation process, hence NOT part of the grammar. Also, in contrast to Ginzburg (2012), who does not employ intentional categories but, nevertheless, requires a one-to-one default mapping between linguistic forms and illocutionary forces, DS does not impose the derivation of an explicitly represented speech-act type for every utterance (see also Sperber and Wilson 1995, p. 244). Instead, speech-act characterisations are optional inferences and in DS-TTR they are implemented as such (see, e.g. Purver et al. 2010; Eshghi et al. (subm) Gregoromichelaki, to appear). This is because it is assumed that the linguistically provided information must be highly underspecified, namely just an indication of sentence mood as, e.g. declarative, interrogative, imperative, so that participants can negotiate and derive the significance of their actions jointly. In DS, such specifications are translatable into semantic features, e.g. inclusion or not of the actual world in the evaluation of truth conditions (see, e.g. Huntley 1984; Farkas 1992; Gregoromichelaki 2006, 2011) or the employment of distinct semantic objects as in Ginzburg (2012), Portner (2004). Following Poesio and Rieser (2010, see Sect. 2), we also have argued for the inclusion of micro-conversational events in the TTR representation (Purver et al. 2010), since these are concrete features of the discourse situation that underpin various (sub-sentential) contextual effects on meaning like shifts of the world of evaluation in metarepresentational cases, conditionals, etc. (Eshghi et al. (subm) Gregoromichelaki 2006, 2011, to appear). Hence, the performance of sub-sentential locutionary acts by participants is recorded incrementally in the grammar model, which is an operation required for purposes like the assignment of referents to indexicals like *I*, *you*,

here and *now*. Such micro-events are essential for the modelling of split utterances, where, as we saw earlier in (7)–(9), such indexicals switch reference mid-sentence.

Further than that, we suggest that the performance of (conventionalised) illocutionary speech acts by various linguistic means (*mood*) is achieved by (subpersonal) use-neutral procedural instructions[9] following the *functions* of such grammatical devices as described in e.g. Millikan (2005, Chaps. 8 and 9). Such functions are conceptualised as reproducible patterns of activity involving the complementary contribution of both speaker and hearer in order to be accomplished. In contrast to the view taken by Ginzburg (2012), in our view, many basic linguistic representations are not inherently differentiated along the descriptive/directive divide (Millikan 2005, Chap. 9), but can become so differentiated through explicit conceptualisation of their function, as it happens in cases of metacommunicative interaction. As argued in Gregoromichelaki (to appear), Eshghi et al. (subm) such functions can be implemented via adopting mechanisms of context update as in Beyssade and Marandin (2006), which modify the commitment record of speaker/hearer after each utterance has been performed. In our view, crucially, such mechanisms do not require, in addition, the involvement of personal, intentional mechanisms attributing mental states to interlocutors in order for their function to be accomplished; in our view, such attributions can be accomplished only derivatively (see also Pickering and Garrod 2004, 2012). In contrast to Ginzburg (2012) who, even though he has set out all the necessary mechanisms for the requisite underspecification, in addition employs default illocutionary-force descriptions, we do not assume that explicit conceptual descriptions of what the participants are doing in the conversation have to be encoded in the common ground in terms of a range of pre-specified speech acts that the speaker, or the grammar, imposes. This hard-wired encoding of speech acts is likely to cause problems for the data that concern us in this chapter. For example, the grammar might assign default assertive force to an utterance; however, a subsequent *why?* question, more plausibly queries, and hence establishes, the actual speech act that has been performed:

Example (10) A: (Let me remind you who is the boss around here!) You leave town tonight. And when you're gone, you stay gone, or you be gone.
B: Right, o.k. (But) why?
['*Why are you ordering me to leave town tonight?*'/ '*Why should I leave town tonight?*'/ #'*Why are you* asserting *I leave town tonight*']

Example (11) [Context: Mary, seeing Peter about to throw a snowball, says threateningly:]
Mary: Yeah, just you dare. Go on. Throw it.
Peter: Why? What are you gonna do?

[9] Whether there are "grammaticised" associations between moods/grammatical devices and speech acts is an empirical issue to be decided on a language-by-language basis.

[# '*Why are you ordering me to throw it?*'/ '*Why are you threatening me?*']
[adapted from Wilson and Sperber 1988]

In our view, the range of actions that can be performed with verbal means is culturally specific (see also Wong, this volume), indefinitely extendible and negotiable; and there is no reason to assume that explicit linguistic or conceptual descriptions can be derived for the precise effect of each utterance, especially if these have to be considered as, even weak, defaults. However, this is not to preclude inferential reasoning about the nature of the speech act where warranted, for example, where participants have conceptualised such actions as in the occurrence of explicit performatives (or other evidence of 'metapragmatic' awareness) or, more implicitly, where trouble arises and inferential procedures have to be employed that require the conceptualisation and description of the discourse situation. Indeed, as the DS formalism is designed to interact with context incrementally at any point, the possibility of deriving action/attitude attribution or planning exists as an optional inferential mechanism. Moreover, crucially, such procedures can be invoked at any sub-sentential point during an utterance, instead of being considered 'root-clause' phenomena. In our view, a processing architecture, the 'grammar', should enable these inferences when the appropriate function of a turn is at issue (e.g. in practices of 'repair' or when one is being held 'accountable' and has to conceptualise what they have been doing), but they are not required for intelligibility or the determination of grammaticality. Such speech-act descriptions are also derivable retrospectively: for example, as a result of an interlocutor's feedback, one can assign a particular force (even to one's own contribution) that had not occurred to them beforehand.

As an illustration, consider that continuations in split utterances, besides being the continuation of the other's utterance, can also perform diverse functions (see also Purver et al. 2009). In (26), B's continuation seems to function as a clarification of A's intended query as well as a continuation of that query, which can be in effect a request for giving back an item that belongs to A. The hearer's response (as e.g. in (27)) will determine whether all, some, or none of these characterisations obtain:

Example (26) A: Did you give me back
B: your loppers? They are there, take them.

Example (27) A: No I meant the secateurs. / I don't want them anymore, I was just asking. / Thanks!

Others have pointed out that continuations can function as, e.g. adjuncts (Fernandez and Ginzburg 2002) or clarification requests (Purver et al. 2003). In all these cases, underspecification of the speech act initiated by a speaker is crucial for deriving the negotiable nature of the actions performed jointly in conversation. Given the sequential context (as described also in Conversation Analysis, see, e.g. Schegloff (2007)), and goals to be fulfilled by the participants (as provided by the procedural analysis of NL input as modelled in DS), multiple speech acts can be performed by the use of a single grammatical construction shared across turns between interlocutors:

Example (28) A: Go away
B: and if I don't [Antecedent of conditional threat; Continuation; Query]
A: I'll smash your face [Consequent of conditional threat; Continuation; Reply; Prediction etc.] [natural data]

Example (29) Freddie (who fancies the boss's daughter): I didn't know you were…
Mike (who goes out with boss' daughter): banging the boss' daughter?
[Completion/Clarification/Assertion (informing)/Challenge/Provocation] [Cemetery Junction]

Notice that these are not just cases of 'one action being the vehicle for another' (or *indirect speech acts*) as identified by, e.g. Levinson (2012) and Schegloff (2007). Here multiple actions are performed during the unfolding of a single propositional unit and, in our view, there is no definitive description that conceptualises what the participants are doing at each sub-sentential point in order to fit it within a range of pre-specified speech-act characterisations (contra Searle's *principle of expressibility,* Searle (1969, p. 18, 1979, p. 134)). Neither is it necessary to assume that first a very general illocutionary force is derived, e.g. assertion, query etc., and then further inferences are drawn to modify or further specify what happens in actual use. Such solutions usually lead to incompatible assignments of forces (see e.g. Asher and Lascarides 2001).

What we have just described shows that in an appropriate sequential environment, co-construction can be employed for the (implicit) performance of speech acts without first establishing propositional contents. However, we have argued, even further than this, that not only propositional/subpropositional contents but also the unarguably sub-personal mechanisms of the grammar itself can be utilised for the performance of speech acts (Gregoromichelaki 2013b; Gregoromichelaki et al. 2013b). Based on the fact that syntax and interpretation are both conceptualised in DS as a single action system, actions in dialogue can be accomplished just by establishing 'syntactic conditional relevances', i.e. exploiting the grammatical dependencies themselves to induce a response by the listener (*grammar-induced speech acts*). In the following, for example, incomplete syntactic dependencies can be initiated by a speaker inviting the hearer to fulfil them, thus forming, e.g. a query–answer pair during the derivation of a single proposition (see also Jaszczolt et al., this volume):

Example (30) A: Thank you mister…
B: Smith, Tremuel [natural data]

Example (31) A: Shall we go to the cinema or…
B: let's stay at home [natural data]

Example (32) A: And you're leaving at…
B: 3.00 o'clock

Example (33) Man: and this is Ida
Joanna: and she was found?
Man: she was found by a woman at Cheltenham. [Catwoman]

Example (34) A: And they ignored the conspirators who were…
B: Geoff Hoon and Patricia Hewitt [radio 4, Today programme, 06/01/10]

Example (35) Hester Collyer: It's for me.

Example (36) Mrs Elton the landlady: And Mr. Page?

Example (37) Hester Collyer: is not my husband. But i would rather you continue to think of me as Mrs. Page. [The Deep Blue Sea]

Example (38) 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]

Example (39) 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 use coal furnace you see…
[BNC, H5H: 59–61]

There is no reason to suppose here that the speakers had a fully formed propositional message to convey before they started production, in fact these formats exactly contradict various assumed direct [speech act ↔ syntax] mappings as in Ginzburg (2012). Moreover, in some contexts, invited completions of another's utterance have been argued to exploit the vagueness/covertness/negotiability of the speech act involved to avoid overt/intrusive elicitation of information:

Example (40) (Lana = client; Ralph = therapist)
Ralph: Your sponsor before…
Lana: 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.
Ralph: A year. Well, I'm not perhaps the expert in this case at all. However, I must admit that you're still young in (.) sobriety and I think that maybe still working with a woman for a while might be
Lana: Yeah
Ralph: in your best interest. [Ferrara 1992]

Here the therapist uses an invited completion in a way that gives the patient the opportunity to assign it the force of a query or not and hence to reveal or not as much information as she is willing to reveal.

As argued in Kempson et al. (2009a, b) and Gregoromichelaki et al. (2011), what is essential in accounting for all these data, along with 'disfluencies' which abound in actual conversation (see earlier example (1)), is an incremental grammar

that models the parallel course and common mechanisms of parsing/production at an appropriate sub-sentential/sub-propositional level. Along with other researchers, we have suggested that intentions/plans should not be seen as causal factors driving coordination but, instead, as discursive constructs that are employed by participants, as part of a (meta)language regarding the coordination process itself, when participants need to conceptualise their own and others' performance for purposes of explicit deliberation or accountability when trouble arises (see Mills and Gregoromichelaki 2010, for experimental evidence). One such device, we argue, are *why*-interrogatives to which we turn next.

4.4 Why?-Interrogatives and the (Split-)Turn-Taking Puzzle

In our view, *why*-interrogatives, even when they appear to simply request the provision of a cause for an event/phenomenon, have most frequently the function of requesting an account for some previous action (as argued also in Bolden and Robinson 2011; Robinson and Bolden 2010). Since, usually, this is a dispreferred option in discourse, this would explain their infrequency and non-locality as established in Ginzburg (2012). In the following, two friends are discussing the name Lea gave to her daughter. She is worried that the English pronunciation "Rachel" (in line 4) might not be acceptable by the religious Jewish community in which they live. The *why?* fragment is used to challenge this attitude:

Example (41)
1. ZIV: What're you calling her.
2. (.)
3. ZIV: You don' kno[w (yet).]
4. LEA: [Ra]chel.
5. (0.8)
6. ZIV: ↑ That's cu::[:te. ↓]
7. LEA: [Mm hm]:,
8. ZIV: .hhh That's cu:te,
9. LEA: I hope. I hope it sticks. Ehhh ((laughter))
10. ZIV: Why:. [*'Why do you* HOPE *it sticks'/ 'Why are you saying you hope it sticks'/ 'Why should it (not) stick' 'Why are you laughing?' etc.*]
11. (.)
12. ZIV: You dec[ided before?]
13. LEA: [I don' know 'cause it's an] English na(h)me.=h
14. ZIV: .hhhh So:,
15. LEA: .hh So you [never know.]
16. ZIV: [You're thuh mothe]r,
17. LEA: Yeah that's true.
18. (0.5)

[Excerpt 5 (CF 4889) from Bolden and Robinson 2011, our comments]

Unlike Ginzburg's ambiguity analysis (see Sect. 3) which seeks to establish a clear-cut distinction between two categories of elliptical *why?* fragments, i.e. querying the reason behind a recently performed speech act versus request for provision of a cause for its propositional content, in many cases as in (41) above, and in (42)–(44) below, it is indeterminate which, if any, of the two is the case:

Example (42)
CAI: Ho:ld on. lemme get my paper, >.h< There was a uh:m (.) .mtch a:rticle in thuh paper about you toda:y,
ROB: .mtch= .hhhhh hhhhh That's not goo:d,
CAI: Why:., ['*Why is it not good?*'/ '*Why are you saying it's not good*']
(1.0)
ROB: 'Cau:se.
(0.2)
CAI: 'Cause what.
ROB: .hhhh I's not.=h
(2.2)
CAI: You don' even know what it sa:id.
ROB: We:ll, (0.3) .hhhh (.) I'm not there to defend myself so hh hh
CAI: It was a good article,
ROB: Eh::=h
CAI: I'll tell you in a second once I fi:nd what=you're lookin' for here.
[Excerpt 6 (CH 6100, adapted from Bolden and Robinson 2011, comments in italics ours]

Example (43)
A: Sorry!
B: Why?

Example (44)
Woman: Salvo?
Salvo: yes?
Woman: always the cop
S: Why? ['Why (are you saying) "always the cop"']
W: you asked all the questions, you told me nothing about yourself
S: what's to tell? [translated from Italian, Inspector Montalbano, Season 3, Episode 2]

In many cases, there is no reason why the interlocutors should be presumed to have to resolve the vagueness involved in such questioning. However, this resolution is grammatically imposed as an unavoidable condition on understanding and response in Ginzburg's bifurcation into two distinct constructions. Moreover, despite the assumed conventional arbitrariness of 'constructions', as can be seen in (44), such phenomena occur crosslinguistically, which indicates rather the involvement of more general mechanisms. In addition, the assumptions underlying the postulation of a separate why_{meta} reading occur freely in other 'constructions', which indicates that the phenomenon needs general treatment:

Example (45) A: Since we're here in the Olympic velodrome, do you fancy another lap?
B (panting): Really? OK then ['*Are you* really *asking me if* I fancy another lap?']

Further evidence pointing in the same direction comes from *why*-ellipsis cases where a linguistic antecedent is absent. Despite the fact that Ginzburg provides rules for *why*-ellipsis resolution that require the presence of linguistically derived content, matching a salient fact accepted in the common ground, *why* fragments can very naturally query salient non-verbal actions or contents that do not directly match the contents introduced in the common ground by some previous utterance:

Example (46) [Context: A comes in the room and punches B]
B: Why?

Example (47) A: this is great!
B: shrugs, winces
A: Why? What is not to like about college?

Example (48) Mary: What is that?!
John: It's a tyre lever.
Mary: Why? [# '*Why is this a tyre lever*'/ # '*Why are you asserting that this is a tyre lever*']
John (nodding towards the house): 'Cause there were loads of smackheads in there, and one of them might need help with a tyre. If there's any trouble, just go. I'll be fine. [Sherlock Holmes, BBC, Season 3, episode 3]

Example (49) A: I feel it's the right thing to do. It's the reason I'm here.
B: Why? To shame us over events best relegated to history?

Example (50) A: I lost you
B: Why, were you following me? [The Quiller Memorandum]

Example (51) Dobri: I gave my life to Janus. But now it's here… …and I… I want to keep my shitty life
Becky: Why? You think he's going to kill you? [Utopia, Channel 4, Series 2, Episode 6]

Example (52) A: I've got a date!
B: Oh GOD!!!
A: Why? What?

Example (53) A: You're from Yorkshire, aren't you?
B: Why?
A: You walk into rooms and sit down in them

Ginzburg's account cannot deal with data like those in (46)–(53) because the grammar is distinguished from general contextually-dependent action. As a result, the lexical entries for *why*-fragments make reference to the presence of a *sign* (i.e. an object with phonological and grammatical features) in order to be able to license the occurrence of *why?*. But no such sign has occurred in (46)–(53). And even though (46)–(47) could be dealt with the *genre-accommodation* operations Ginzburg defines to coerce a proposition in the common ground, the constructional articulation of the grammar prevents the data from actually being able to be so handled: under the constructional approach, which makes reference to linguistic 'signs', they could not be handled in principle because it does not sound plausible that the accommodated proposition would need to be accompanied with the full phonological/syntactic features of a potential utterance that never occurred. In contrast, according to our action-oriented DS approach, the grammar is not concerned with defining linguistically pre-specified context conditions for such resolutions. This is because in all cases of language use, the context is constructed on the fly by the interlocutors themselves, so all conceptualisations and articulations of content, even the most mundane lexical choices, once constructed, put forward assumptions as 'presupposed'. Therefore, the grammatical specifications need to be able to facilitate such meshing with non-verbal actions, the material circumstances, and inferentially-derived contents without the need to distinguish 'accommodation' techniques for 'exceptional' uses. What is presented as 'accommodation', according to us, is the usual case in conversation. Accordingly, because DS eschews a separate syntactic level of licensing, it defines morphosyntactic constraints with reference to semantic properties. Hence, there is no problem arising as to how 'accommodated' propositions can be the source of coherence for fragmental sub-sentential elements (for the same point regarding ellipsis in general, see Gregoromichelaki 2012; Kempson et al. to appear).

The same approach can be taken, it seems to us, for the main claim in Ginzburg's analysis of the TTP. His description of this phenomenon relies on data discussed in Sect. 3 earlier (see examples (21)–(24)) that certain readings are unavailable according to whether the user of the *why*?-fragment was previously the speaker or not. These data are repeated below, summarised in a simplified manner:

Example (54) A: Which members of our team own a parakeet?
B: Why?
(a) why_{direct} reading: # '*Why own a parakeet*?'
(b) why_{meta} reading: '*Why are you asking which members of our team own a parakeet?*'

Example (55) A: Which members of our team own a parakeet?
A: Why?
(a) why_{direct} reading: '*Why own a parakeet?*'
(b) why_{meta} reading: # '*Why am asking this?*'

Example (56) A: Which members of our team own a parakeet? Why am I asking this question?

First of all, we contend that the alleged missing readings are not impossible, especially when the queried previous speech act is not itself a query:

Example (57) A: Piss off. Why? Probably because I hate your guts.
Example (58) Stop it! Why? Because I'm your boss, that's why!
Example (59) Careful! Why? Because you're clumsy that's why
Example (60) Fuck off! You know why? 'Cause none of you got the guts to be what you want to be.
Example (61) (In March) Merry Christmas! Why? Because I feel festive!
Example (62) [public prayer] God, thank You for my suffering. Why? Because I am being perfected in it.
Example (63) Davidson should have used more epitaphs. Epitaphs? Why?… Sorry, I meant epithets.
Example (64) Bo came to the party. Bo? Why on earth Bo now? Freudian slip, sorry.

But even with queries, as shown in (65), in multiparty dialogues, where the accountability for the speech act just performed is manifestly joint, these readings can occur:[10]

Example (65) Mum to Dad: Ask your daughter where she was tonight.
Dad to Daughter: Where did you go tonight? <Turning to Mum> Why? [= *'Why are we asking her this?'*] What has she done now?

In our view, even though some such readings have the flavour of 'rhetorical questions' (i.e. queries for which it is not really implied that the speaker does not know the answer, or where the speaker does not request information from the hearer), it seems to us that, e.g. (57) does not require such a reading any more than does (56), Ginzburg's example showing the viability of the alternative reading when a full *why*- interrogative is used. Moreover, if we assume that in general the reason motivating a speaker's utterance cannot be an issue salient enough in the context, how do we explain the implicit propositional argument of *because* explanations following perfectly felicitously, and with high frequency, any type of utterance:

Example (66) Are you in fact going to the funeral? Because I heard you won't.

In fact, one can claim that it is the aptitude of such follow-up explanations that makes why_{meta} queries redundant in most cases, not any inherent grammatical constraint.

But further than this, when the TTP test is applied to split utterances, it appears that distinct empirical results are obtained: given a turn posing a query but split between two interlocutors, the possible interpretations of a subsequent *why?* depend not only on the most recent SPEAKER but also on who can be taken as the agent ac-

[10] Multi-party dialogue data that, in our view, indicate similar results were also noted in Ginzburg (1998) but were taken as leading to distinct conclusions.

countable for the speech act performed, which, in these cases, might be a role distinct from the notion of 'speaker' that is tracked by indexical pronouns like *I* and *my*:

Example (67) A to C:Have you finished sharpening (his)…
B to C/A: my loppers?
B to A: Why? ['*Why are you asking C whether C has finished sharpening B's loppers?*']
A to B: Because I want her to sharpen my secateurs too.

Such data, what we have called the (Split-)Turn Taking Puzzle (STTP), are beyond the explanation of the TTP in Ginzburg (1997, 2012) because, in our view, Ginzburg's grammar does not incorporate a notion of incrementality, with context updates at each word-by-word stage, as a fundamental feature in the architecture of the model.

In order to integrate optional speech-act information, in Purver et al. (2010), we have assumed that the DS apparatus manipulates representations in the TTR language. This is because TTR provides a multidimensional representational format with a well-defined semantics as developed in Cooper (2005, 2012) and Ginzburg (2012). Moreover, through its notion of subtyping, TTR allows the manipulation in the grammar of underspecified objects, through partially specified types, which can be progressively specified/instantiated as more information becomes available. As a result, it becomes possible to articulate highly structured models of context, where uniform representations of multiple types of information can be supplied and their interaction modelled (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 a set of actions, 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 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 subpropositional record types, so that TTR is particularly well-suited for representing how partial semantic information is stepwisely 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 (see Gregoromichelaki and Kempson, to appear). In our view, the latter is what is needed for making explicit the contents that are required in the resolution of fragmentary *why?*-interrogatives, and our explanation of STTP, consequently, also relies on such representations.

Like Ginzburg (2012), our explanation of the STTP puzzle takes the why_{meta} interpretation as querying the intention/plan behind an agent's speech act (locutionary or illocutionary). Following Poesio and Rieser (2010), in Purver et al. (2010), we have suggested that each word utterance induces the context record to be augmented with the inclusion of an event (a 'micro-conversational event' in Poesio and Rieser's

terminology). Such event descriptions include discourse situation and participant information as well as who is uttering this particular word. This is the information that is standardly needed to account for the resolution of indexicals like *I*, *you*, *here,* etc. However, even though the grammar records who the utterer (the agent of the locutionary act) is for such purposes, note that our approach does not necessitate that illocutionary force and therefore intention/commitment information are available by default *prior* to the processing of an utterance such as a *why*-interrogative: instead, seeking to interpret such queries can be the trigger for optional (speech-act representation) rules to apply. Hence, this approach is perfectly compatible with the general view on 'intentions' as post-factum discursive constructs (see e.g. Suchman 2007) and the fact that conversational participants can negotiate the content of their speech acts, with speech-act assignments able to emerge retrospectively. Since, in our view, ellipsis resolution requires the potential for immediate representation of a salient feature of the context, the infelicity of the reading in (55) shows that, in these cases, the speaker's plan behind their speech act is, in general, not a parameter salient enough that hearers need to consider to ground the utterance (as indeed Ginzburg (2012) notes). In contrast, what a speaker does, in terms of micro-conversational events (or indeed non-verbal actions) can be salient enough, as (46)–(47) show. If we assume such an explanation and consider the data in (65) and (67), the TTP then relates to who can be held accountable for performing the relevant act, and hence can be asked to justify their actions. However, unlike Ginzburg (2012), we do not wish to grammaticise such a fact, since, as the STTP, (67), and other cases show, the folk notion of 'speaker' as mapping directly to the role of the agent of the speech act is not adequate. Ginzburg himself contends with similar problems concerning the concept of 'hearer' in multiparty dialogue. And, as shown in Levinson (1988), the decomposition of the concepts of 'speaker' and 'hearer' for various purposes needs to be allowed freely as an option in the grammar.

For our concerns with the (S)TTP, in terms of Goffman's (1981) distinctions among 'speaker' roles, the relevant agent is the 'principal'. In (67), the utterer (the 'animator') of a completion (the final speaker in the general sense, and as indexed by pronouns like *my*) can felicitously ask elliptical why_{meta}-questions of the original utterer, because although B's fragment *my loppers?* completes A's question, B does not necessarily assume responsibility for the performance of the illocutionary speech act. In fact, it is B's *why?* question that can establish this fact in the common ground: its use shows to A and C that A is solely accountable for the query to C as B dissociates explicitly from it (it can be the same in (65) too). Now, A must be taken as the agent accountable for the querying speech act even though there is a sequence of utterance micro-events which A and B have performed severally to accomplish it. The availability of the why_{meta} reading then follows, even though apparently in contrast to (55). In some cases, then, even though the turn is collaboratively constructed, the original speaker maintains the accountability for the turn even though it was completed by somebody else. In other cases, see e.g. (13)–(14) earlier, this is not the case: the eventual content derived has to be taken as solely attributable to the second speaker. And, in even other cases, e.g. in (7)–(9) earlier, this is indeterminate and not relevant to the processing of the dialogue. Hence, in our view, there is no reason for such specifications to be encoded necessarily as they would prevent an

account of continuations as such, i.e. as continuing what somebody else has so far offered, but making use of it perhaps in a new way.

In sum, we claim that the view emerging from such 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. Within this model, '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 in Poesio and Rieser 2010). In the exercise of their grammatical knowledge in interaction, participants justify Wittgenstein's view that 'understanding is knowing how to go on' (Wittgenstein 1980), even on the basis of sub-propositional, sub-sentential input with no reasoning intervening. 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 if required. 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 conceptualised, does not imply that they operate in the background when participants engage in (unconscious, subpersonal) practices that can be described from the outside in explicit propositional terms. 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, we believe that conversational participants manifest their ability to 'make explicit' the practices afforded to them implicitly by subpersonal procedures either when communication breaks down or when they need to verbalise/conceptualise the significance of their actions (for a similar account of practices at other higher levels of coordination, see Piwek 2011).

5 Conclusion

In conclusion, dialogue phenomena like fragmentary and split utterances are not phenomena of performance dysfluency but a diagnostic of essential properties of NL know-how. The problem standard syntactic theories have in dealing with dialogue data can be traced to the assumption that it is sentential strings with propositional interpretations that constitute the output of the grammar, along with the attendant methodological principle debarring any attribute of performance within the grammar-internal characterisation. We have argued here that such phenomena cannot be handled without radically modifying the competence-performance distinction as standardly drawn, even more radically in our view than the significant steps in this direction that Poesio and Rieser (2010), Ginzburg (2012), and others (see e.g. Peldszus and Schlangen 2012; Schlangen 2003) have already taken. We believe that the competence/performance methodology, 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 NL, 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 1984; McDowell 1998).

In the domain of semantics, work by Ginzburg (2012), Cooper (2012), and Larsson (2011), among others, presents a significant advance in that it does not restrict itself to the modelling of informational discourse but, instead, attempts to describe the fine-grained structure of conversational exchanges, explores the ontologies required in order to define how speech events cause changes in the mental states of dialogue participants (see also Poesio and Rieser 2010), and attempts to integrate perception and semantic conceptualisation in a unified framework (TTR). But, following standard assumptions, these models also define syntax independently (Poesio and Rieser 2010) and statically (Ginzburg 2012), which, in our view, prevents the modelling of the fine-grained incrementality observable in the split-utterance data. We believe that what is needed is a domain-general action-oriented model that accounts for both the sub-sentential, supra-sentential and cross-modal structure of an interaction (a *grammar*). This revision of what kind of knowledge a grammar encapsulates changes the view of the semantic landscape. The instrumentalist Davidsonian stance towards the content assigned to sub-sentential constituents, as subordinate to sentential contents, needs to be revised in that sub-sentential contributions provide the locus for as much, and as significant (externalised) 'inference' and coordination among participants, as any propositional contributions. From this perspective, the full array of dialogue data demands a grammar-internal characterisation, in that the licensing of the complete structure, and ultimately the discourse effects of such moves, relies on syntactic/semantic constraints. As a result, in our view, a uniform account of such data within the grammar itself can only be given with a shift of perspective into one in which NL knowledge is seen as action-based (procedural), i.e. a set of unencapsulated processing mechanisms.

References

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

Arundale, Robert B. 2008. Against (Gricean) intentions at the heart of human interaction. *Intercultural Pragmatics* 5 (2): 229–258.

Asher, Nicholas, and Alex Lascarides. 2001. Indirect speech acts. *Synthese* 128 (1–2): 183–228.

Beyssade, Claire, and J.-M. Marandin. 2006. *From complex to simple speech Acts: A bidimensional analysis of illocutionary forces*. Potsdam: Brandial.

Blakemore, Diane. 1987. *Semantic constraints on relevance*. Oxford: Wiley-Blackwell.

Bolden, Galina B., and Jeffrey D. Robinson. 2011. Soliciting accounts with why-interrogatives in conversation. *Journal of Communication* 61 (1): 94–119.

Brandom, Robert B. 1994. *Making it explicit: Reasoning, representing, and discursive commitment*. Cambridge: Harvard University Press.

Bratman, Michael E. 1990. What is intention? In *Intentions in communication*, ed. Philip R. Cohen, Jerry Morgan, and Martha E. Pollack. Cambridge: MIT Press.

Bratman, Michael E. 1992. Shared cooperative activity. *Philosophical Review* 101:327–341.

Bratman, Michael E. 1993. Shared intention. *Ethics* 104:97–113.
Barwise, Jon, and John Perry. 1983. *Situations and attitudes*. Cambridge: MIT Press.
Cann, Ronnie, Ruth Kempson, and Lutz Marten. 2005. *The dynamics of language*. Oxford: Elsevier.
Chomsky, Noam. 2000. *New horizons in the study of language and mind*. Cambridge: Cambridge University Press.
Clark, Herbert H. 1996. *Using language*. Cambridge: Cambridge University Press.
Clark, Herbert H., and Jean E. Fox Tree. 2002. Using uh and um in spontaneous speech. *Cognition* 84:73–111.
Cooper, Robin. 2005. Records and record types in semantic theory. *Journal of Logic and Computation* 15 (2): 99–112.
Cooper, Robin. 2012. Type theory and semantics in flux. In *Philosophy of linguistics* (part of the Handbooks of the Philosophy of Science series), eds. R. Kempson, T. Fernando, and N. Asher, 14, 271–323.
Cooper, Robin, and Aarne Ranta. 2008. Natural languages as collections of resources. In *Language in flux*, eds. Robin Cooper and Ruth Kempson, 109–120. London: College Publications.
Davidson, Donald. 1967. Truth and meaning. *Synthese* 17:304–323.
Davidson, Donald. 1986. A nice derangement of epitaphs. In *Truth and interpretation*, ed. E. Lepore, 433–446.
Davidson, Donald. 1994. The social aspect of language. In *The philosophy of Michael Dummet*, eds. B. McGuiness and G. Oliveri. Oxford: Blackwell.
De Ruiter, Jan-Peter, Holger Mitterer, and Nicholas J. Enfield. 2006. Projecting the end of a speaker's turn: A cognitive cornerstone of conversation. *Language* 82 (3): 515–535.
Eshghi, Arash, Patrick G. T. Healey, Matthew Purver, Christine Howes, Eleni Gregoromichelaki, and Ruth Kempson. 2010. Incremental Turn Processing in Dialogue. In Proceedings of the 16th Annual Conference on Architectures and Mechanisms for Language Processing (AmLAP), York, UK, September 2010.
Eshghi, Arash, Michael Purver, and Julian Hough. 2011. *Dylan: Parser for dynamic syntax. Technical report*. London: Queen Mary University of London.
Eshghi, Arash, J. Hough, Michael Purver, Ruth Kempson, and Eleni Gregoromichelaki. 2012. Conversational interactions: Capturing dialogue dynamics. In From *Quantification to conversation*, eds. L. Borin and S. Larsson. London: College Publications.
Farkas, Donka. 1992. On the semantic of subjunctive complements. In *Romance languages and modern linguistic theory*, ed. P. Hirschbühler. Amsterdam: John Benjamins.
Fernandez, Raquel, and Jonathan Ginzburg. 2002. Nonsentential utterances: A corpus study. Traitement automatique des langages. *Dialogue* 43 (2): 13–42.
Ferrara, Kathleen. 1992. The interactive achievement of a sentence: Joint productions in therapeutic discourse. *Discourse Processes* 15:207–228.
Frege, Gottlob. 1884/1980. *The foundations of arithmetic* (Trans. J. L. Austin). Second Revised Edition. Evanston: Northwestern University Press.
Gargett, Andrew, Eleni Gregoromichelaki, Christine Howes, and Yo Sato. 2008. Dialogue-grammar correspondence in Dynamic Syntax. In Proceedings of the 12th SemDial (LonDial).
Gargett, Andrew, Eleni Gregoromichelaki, Ruth Kempson, Matthew Purver, and Yo Sato. 2009. Grammar resources for modelling dialogue dynamically. *Journal of Cognitive Neurodynamics* 3 (4).
Ginzburg, Jonathan. 1997. On some semantic consequences of turn taking. In Proceedings of the 11th Amsterdam Colloquium, eds. Paul Dekker, Martin Stokhof, and Yde Venema. Institute for Logic, Language, and Computation, University of Amsterdam.
Ginzburg, Jonathan. 1998. *Uttering in dialogue*. Ms. Hebrew University of Jerusalem.
Ginzburg, Jonathan. 2012. *The interactive stance: Meaning for conversation*. Oxford: Oxford University Press.
Ginzburg, Jonathan, and Robin Cooper. 2004. Clarification, ellipsis, and the nature of contextual updates in dialogue. *Linguistics and Philosophy* 27 (3): 297–365.
Goffman, Erving. 1981. *Forms of talk*. Philadelphia: University of Pennsylvania Press.

Goodwin, Charles. 1979. The interactive construction of a sentence in natural conversation. In *Everyday language: Studies in ethnomethodology*, ed. G. Psathas, 97–121. New York: Irvington Publishers.

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

Goodwin, Charles. 1995. Co-constructing meaning in conversations with an aphasic man. *Research on Language and Social Interaction* 28 (3): 233–260.

Gregoromichelaki, Eleni. 2006. *Conditionals in Dynamic Syntax*. PhD thesis, Oxford University. London: King's College.

Gregoromichelaki, Eleni. 2011. Conditionals in dynamic syntax. In *The dynamics of lexical interfaces*, ed. R. Kempson, E. Gregoromichelaki, and C. Howes, CSLI Publications.

Gregoromichelaki, E. 2012. Review of J. Ginzburg (2012) *The Interactive Stance*. *Folia Linguistica* 47 (1): 293–316.

Gregoromichelaki, Eleni. 2013a. Clitic left dislocation and clitic doubling: A dynamic perspective on left-right asymmetries in Greek. In *Rightward movement in a comparative perspective*, eds. Gert Webelhuth, Manfred Sailer, and Heike Walker. John Benjamins.

Gregoromichelaki, Eleni. 2013b. Grammar as action in language and music. In *Language, Music and Interaction*, eds. M. Orwin, C. Howes, and R. Kempson. London: College Publications.

Gregoromichelaki, Eleni, and Ruth Kempson. to appear. Joint utterances and indirect reports. In *The Pragmatics of Indirect Reports,* ed. Alessandro Capone, Ferenc Kiefer, and Franco Lo Piparo. Dordrecht: Springer.

Gregoromichelaki, Eleni, and Ruth Kempson 2013. The role of intentions in dialogue processing. In *Perspectives on linguistic pragmatics (Perspectives in Pragmatics, Philosophy & Psychology*, Vol. 2, ed. A. Capone, F. Lo Piparo, and M. Carapezza). Dordrecht: Springer.

Gregoromichelaki, Eleni, Yo Sato, Ruth Kempson, Andrew Gargett, and Christine Howes. 2009. Dialogue Modelling and the Remit of Core Grammar. In: *Proceedings of IWCS* 2009.

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

Gregoromichelaki, Eleni, Ruth Kempson, and Ronnie Cann. 2012. Language as tools for interaction: Grammar and the dynamics of ellipsis resolution. *The Linguistic Review* 29 (4): 563–584.

Gregoromichelaki, Eleni, Ronnie Cann, and Ruth Kempson. 2013a. Coordination in dialogue: Subsentential speech and its implications. In *Brevity*, ed. L. Goldstein. Oxford University Press.

Gregoromichelaki, E., Ruth Kempson, Christine Howes, and Arash Eshghi. 2013b. On making syntax dynamic: The challenge of compound utterances and the architecture of the grammar. In *Alignment in Communication: Towards a New Theory of Communication*, eds. Ipke Wachsmuth, Jan de Ruiter, Petra Jaecks, and Stefan Kopp. Amsterdam: John Benjamins.

Groenendijk, Jeroen and Martin Stokhof. 1991. Dynamic Predicate Logic. *Linguistics and Philosophy* 14(1).

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

Haugh, Michael and Katarzyna M. Jaszczolt. in press. Speaker intentions and intentionality. In *Cambridge handbook of pragmatics*, eds. Katarzyna M. Jaszczolt and Keith Allan. Cambridge: Cambridge University Press.

Hawkins, J. 2004. *Efficiency and complexity in grammars*. Cambridge: Cambridge University Press.

Healey, Patrick G. T. 2008. Interactive misalignment: The role of repair in the development of group sub-languages. In *Language in Flux*, eds. Robin Cooper and Ruth Kempson. London: College Publications.

Howes, Christine, Patrick G. T. Healey, and Gregory J. Mills. 2009. A: An experimental investigation into B:…split utterances In Proceedings of SIGDIAL 2009: The 10th Annual Meeting of the Special Interest Group in Discourse and Dialogue, Queen Mary University of London, Sept. 2009.

Howes, Christine, Matthew Purver, Patrick G. T. Healey, and Gregory J. Mills. 2011. On incrementality in dialogue: Evidence from compound contributions. *Dialogue and Discourse* 2 (1). Special Issue on Incremental Processing in Dialogue.

Huntley, M. 1984. The semantics of the English imperative. *Linguistics and Philosophy* 7:103–33.

Jaszczolt, Katarzyna M. 2005. *Default semantics: Foundations of a compositional theory of acts of communication*. Oxford: Oxford University Press.

Kempson, Ruth, Wilfried Meyer-Viol, and Dov Gabbay. 2001. *Dynamic syntax: The flow of language understanding*. Oxford: Blackwell.

Kempson, Ruth, Eleni Gregoromichelaki, and Yo Sato. 2009a. Incrementality, speaker/hearer switching and the disambiguation challenge. In Proceedings of European Association for Computational Linguistics (EACL Athens 2009).

Kempson, Ruth, Eleni Gregoromichelaki, Matthew Purver, Gregory J. Mills, Andrew Gargett, and Christine Howes. 2009b. How mechanistic can accounts of interaction be? In: Proceedings of Diaholmia, the 13th Workshop on the semantics and pragmatics of dialogue.

Kempson, Ruth, Eleni Gregoromichelaki, Wilfried Meyer-Viol, Matthew Purver, Graham White, and Ronnie Cann. 2011a. Natural-language Syntax as Procedures for Interpretation: The dynamics of ellipsis construal. In Proceedings of the PRELUDE Workshop on Games, Dialogues and Interaction. Lecture Notes in Computer Science. Berlin: Springer.

Kempson, Ruth, Eleni Gregoromichelaki, and Christine Howes. 2011b. *The dynamics of lexical interfaces*. Stanford: CSLI Publications.

Kempson, Ruth, Eleni Gregoromichelaki, and S. Chatzikyriakidis. 2012b. Joint utterances in Greek: their implications for linguistic modelling. In: Proceedings of 33rd Annual Linguistics Meeting "Syntactic Theories and the Syntax of Greek". Thessaloniki, 26–27 April 2012.

Kempson, Ruth, Eleni Gregoromichelaki, and Ronnie Cann. 2012. Context and compositionality: The challenge of conversational dialogue. In *Philosophical and formal approaches to linguistic analysis,* ed. Piotr Stalmaszczyk, 215–240. Berlin: Walter de Gruyter.

Kempson, Ruth, Eleni Gregoromichelaki, Arash Eshghi, and Julian Hough. (to appear). Ellipsis in Dynamic Syntax. In *The Oxford Handbook of Ellipsis*, eds. Jeroen van Craenenbroeck and Tanja Temmerman. Oxford University Press.

Larson, R., and G. Segal. 1995. *Knowledge of meaning: An introduction to semantic theory*. Cambridge: MIT Press.

Larsson, Staffan. 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), 140–148, September 2011.

Larsson, Staffan, and David R. Traum. Information state and dialogue management in the TRINDI dialogue move engine toolkit. *Natural language engineering* 6 (3-4): 323–340.

Lerner, Gene H. 1991. On the syntax of sentences-in-progress. *Language in Society* 20:441–458.

Levinson, Stephen C. 1988. Putting linguistics on a proper footing. explorations in Goffman's concepts of participation. In *Goffman. Exploring the interaction order,* eds. P. Drew and A. Wootton, 161–227. Oxford: Polity Press.

Levinson, Stephen C. 1995. Interactional biases in human thinking. In *Social intelligence and interaction*, ed. Esther N. Goody, 221–260. Cambridge: Cambridge University Press.

Levinson, Stephen C. 2000. *Presumptive meanings*: *The theory of generalized conversational implicature*. Cambridge: MIT Press.

Levinson, Stephen C. 2012. Action formation and ascription. In *The handbook of conversation analysis,* eds. J. Sidnell and T. Stivers. Chichester: Wiley. doi:10.1002/9781118325001.ch6.

Lombardo, Vincenzo, and Patrick Sturt. 2002. Incrementality and lexicalism: A treebank study. In *The lexical basis of sentence processing: formal, computational and experimental issues*, ed. S. Stevenson and P. Merlo, 137–155. Philadelphia: John Benjamins.

McDowell, J. 1998. *Mind, value and reality*. Cambridge: Harvard University Press.

Millikan, Ruth G. 2005. *Language: A biological model*. Oxford: Oxford University Press.

Mills, Gregory J., and Eleni Gregoromichelaki. 2010. Establishing coherence in dialogue: Sequentiality, intentions and negotiation. In Proceedings of the 14th SemDial, PozDial.

Montague, R. 1970. English as a formal language. In *Linguaggi nella Societa et nella Technica,* ed. B. Visentini et al., 188–221. Milan: Edizioni di Communità. Reprinted in Thomason (ed.) 1974, pp. 188–221.

Morgan, Jerry. 1973. Sentence fragments and the notion 'sentence'. *Issues in Linguistics* 719–751.

Morgan, Jerry L. 1973. Sentence Fragments and the Notion 'Sentence'. In *Issues in Linguistics: Papers in Honor of Henry and Renée Kahane,* Eds. Braj Kachru, Robert Lees, Yakov Malkiel, Angelina Pietrangeli, and Sol Saporta, 719–751. Urbana: University of Illinois Press.

Peldszus, Andreas, and David Schlangen. 2012. Incremental construction of robust but deep semantic representations for use in responsive dialogue systems. In *Proceedings of the Workshop Advances in discourse analysis and its computational aspects at Coling 2012, Mumbai, India*, ed. Eva Hajičová.

Phillips, Colin. 2003. Linear order and constituency. *Linguistic Inquiry* 34(1): 37–90.

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

Pickering, Martin J., and Simon Garrod. 2012. 'An integrated theory of language production and comprehension'. *Behavioral and Brain Sciences*.

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

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

Poesio, Massimo, and David Traum. 1998. Towards an axiomatization of dialogue acts. Proceedings of the Twente Workshop on the Formal Semantics and Pragmatics of Dialogues (13th Twente Workshop on Language Technology).

Poesio, Massimo, and Hannes Rieser. 2010. Completions, coordination, and alignment in dialogue. *Dialogue and Discourse* 1 (1): 1–89.

Portner, P. 2004. The semantics of imperatives within a theory of clause types. In *Proceedings of SALT 14*, ed. K. Watanabe and R. B. Young, 235–252. Ithaca: CLC Publications.

Purver, Matthew, Patrick Healey, James King, Jonathan Ginzburg, and Gregory Mills. 2003. Answering clarification questions. In Proceedings of the 4th SIGdial Workshop on Discourse and Dialogue, 23–33, Association for Computational Linguistics, Sapporo, Japan, July 2003.

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

Purver, Matthew, Christine Howes, Eleni Gregoromichelaki, and Patrick G. T. Healey. 2009. Split utterances in dialogue: A corpus study. In: Proceedings of SigDial London, September 2009.

Purver, Matthew, Eleni Gregoromichelaki, Wilfried Meyer-Viol, and Ronnie Cann. 2010. Splitting the I's and Crossing the You's: Context, Speech Acts and Grammar. In SemDial 2010 (PozDial), Poznań, Poland, June 2010.

Purver, Matthew, Arash Eshghi, and Julian Hough. 2011. Incremental semantic construction in a dialogue system. In *9th International Conference on Computational Semantics (IWCS)*, Oxford, January 2011.

Robinson, Jeffrey D., and Galina B. Bolden. 2010. Preference organization of sequence-initiating actions: The case of explicit account solicitations. *Discourse Studies* 12 (4): 501–533.

Sacks, Harvey, Emanuel A. Schegloff, and Gail Jefferson. 1974. A simplest systematics for the organization of turn-taking for conversation. *Language* 50:696–735.

Sacks, Harvey. 1992. *Lectures on conversation*. Vols I and II (G. Jefferson, ed.). Oxford: Blackwell.

Schegloff, Emanuel A. 2007. *Sequence organization in interaction: A primer in conversation analysis*. Cambridge: Cambridge University Press.

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

Schober, Michael F., and Herbert H. Clark. 1989. Understanding by addressees and overhearers. *Cognitive Psychology* 21:211–232.

Searle, John R. 1969. *Speech acts*. *An essay in the philosophy of language*. Cambridge: Cambridge University Press.

Searle, John R. 1979. *Expression and meaning*. Cambridge: Cambridge University Press.

Sperber, Dan and Deirdre Wilson. 1995. *Relevance: Communication and cognition*, 2nd edn. Oxford: Blackwell.

Stainton, Robert. 2006. *Words and thoughts: Subsentences, ellipsis, and the philosophy of language*. Oxford: Oxford University Press.

Stalnaker, Robert. 1979. Assertion. In *Syntax and Semantics 9,* ed. Peter Cole. New York: Academic.

Stone, Matthew. 2004. Intention, interpretation and the computational structure of language. *Cognitive Science* 28 (5): 781–809.

Sturt, Patrick, and Vincenzo Lombardo. 2005. Processing coordinate structures: Incrementality and connectedness. *Cognitive Science* 29:291–305.

Suchman, Lucy A. 1987/2007. *Plans and situated actions: The problem of human-machine communication*. Cambridge University Press.

Wilson, Deirdre, and Dan Sperber. 1988. Mood and the analysis of non-declarative sentences. In *Human agency: Language, duty and value*, ed. J. Dancy, J. Moravcsik, and C. Taylor, 77–101. Stanford: Stanford University Press. Reprinted in A. Kasher (ed.). 1998. *Pragmatics: Critical concepts*, vol. II. Routledge, London: 262–289.

Wittgenstein, Ludwig. 1980. *Remarks on the philosophy of psychology*. vol. 1. Oxford: Blackwell.