Dialogue and Discourse 2(1) (2011) 199–233 doi: 10.5087/dad.2011.109

Incrementality and Intention-recognition in Utterance Processing

Eleni Gregoromichelaki and **Ruth Kempson** {ELENI.GREGOR, RUTH.KEMPSON}@KCL.AC.UK
Department of Philosophy
King's College London,
Strand, London WC2R 2LS, UK

Matthew Purver MPURVER@DCS.QMUL.AC.UK
Department of Computer Science
Queen Mary University of London,
Mile End Road, London E1 4NS, UK

Gregory J. Mills GJMILLS@STANFORD.EDU
Department of Psychology
Stanford University,
450 Serra Mall, Stanford, California 94305, USA

Ronnie Cann R.CANN@ED.AC.UK
School of Philosophy, Psychology and Language Sciences
University of Edinburgh,
Dugald Stewart Building, 3 Charles Street Edinburgh, EH8 9AD, UK

Wilfried Meyer-Viol WILFRIED.MEYER_VIOL@KCL.AC.UK
Department of Philosophy
King's College London,
Strand, London WC2R 2LS, UK

Patrick G. T. Healey PH@DCS.QMUL.AC.UK
Department of Computer Science
Queen Mary University of London,
Mile End Road, London E1 4NS, UK

Editor: Hannes Rieser and David Schlangen

Abstract

Ever since dialogue modelling first developed relative to broadly Gricean assumptions about utterance interpretation (Clark, 1996), it has remained an open question whether the full complexity of higher-order intention computation is made use of in everyday conversation. In this paper we examine the phenomenon of *split utterances*, from the perspective of *Dynamic Syntax*, to further probe the necessity of full intention recognition/formation in communication: we do so by exploring the extent to which the interactive coordination of dialogue exchange can be seen as emergent from low-level mechanisms of language processing, without needing representation by interlocutors of each other's mental states, or fully developed intentions as regards messages to be conveyed. We thus illustrate how many dialogue phenomena can be seen as direct consequences of the grammar architecture, as long as this is presented within an *incremental*, *goal-directed/predictive* model.

Keywords: Dialogue, Gricean accounts of communication, Higher-order intention recognition, Incrementality, Intentions, Mind-reading, Plans, Predictivity, Split-Utterances

Submitted 1/2010; Accepted 1/2011; Published online 5/2011

1. Introduction: Rethinking intentionalism in communication

Ever since the first attempts at modelling communication relative to broadly Gricean assumptions, it has remained an open question what level of complexity of higher-order intention computation is made use of in everyday conversation. In this paper, we argue that the interactive coordination of dialogue exchange can be seen as emergent from the mechanisms of language processing, without either needing representation by interlocutors of each other's mental states or fully developed intentions as regards messages to be conveyed. This conclusion is controversial in that it is not commensurate with a broad swathe of recent pragmatic theorising. Higher-order intention recognition forms the underpinning to Grice's account of non-natural meaning (meaning$_{NN}$) and the subsequent communication models that have been based on it. Central to all such accounts is the assumption that understanding by a hearer involves recognition of the particular proposition a speaker intended to express, via their recognition of that intention. Though the conceptual and psychological problems higher-order intention recognition gives rise to are wellknown, responses to these criticisms have been muted. Either the problems are ignored altogether; or they have resulted in unsubstantiated weakenings of the stringent requirements such recognition places on the recovery of meaning in communication even though purportedly retaining the central tenets of the Gricean paradigm. In this paper, having introduced general philosophical and psychological re-evaluations of the status of higher-order intention recognition, we turn to an additional consideration, the problems raised for Gricean views by the phenomenon of *incrementality* in both comprehension and production as manifested in conversational dialogue. Our particular focus is the phenomenon of so-called *split utterances*, commonly seen in dialogue, in which speakers and hearers reverse roles mid-utterance. To deal effectively with the analysis of such shared productions, we turn to a model in which incrementality is a core property of the grammar formalism. Under this assumption, we first show how, with "syntax" re-defined to be the incremental and monotonic growth of semantic representation, the split utterance phenomenon is straightforwardly both predictable and explainable. We then argue that, relative to this model, recognition of the content of speaker intentions is not a necessary condition for human interaction. Hence, we will conclude, it is not an intrinsic property of communication.

1.1 Intention recognition in communication and dialogue

Grice's account of communication (published as Grice, 1975), based on the notion of "meaning$_{NN}$", was the point of departure for many subsequent pragmatic models (see Levinson, 1983; Bach, 1997; Bach and Harnish, 1982; Cohen et al., 1990, a.o.).[1] It characterised communication as essentially involving rationality and cooperation, displayed by the requirement that cooperative interlocutors must be guided by reasoning about mental states: speaker's meaning, whose recovery is elevated as the fundamental criterion for successful communication, involves the speaker at least (a) having the intention of producing a response (e.g. belief) in the addressee (i.e. having a thought about the addressee's thoughts) and (b) also having a second order intention regarding the addressee's belief about the speaker's second order thought (in order to capture the presumed fulfillment of the communicative intention by means of its recognition). Under this definition, speakers have (at least) fourth order thoughts and hearers must recover speaker's meaning through reasoning about these

1. Note that our arguments here do not necessarily concern Grice's philosophical account, in so far as it is seen by some as just normative, but its employment in subsequent (psychological/computational) models of communication/pragmatics.

thoughts. Early on, philosophers like Strawson (1964) and Schiffer (1972) severally presented scenarios where the criterion of higher-order intention recognition was satisfied even though this still was not sufficient for the cases to be characterised as instances of "communication" (as opposed to covert manipulation, sneaky intentions etc.). This led to the postulation of successively higher levels of intention recognition as a prerequisite for communication, and an attendant concept of "mutual knowledge" of speaker's intentions, both of which were recognised as facing a charge of infinite regress (see e.g. Sperber and Wilson, 1995, 256-77). Although in applications of this account in psychological implementations it is not necessary to assume that explicit reasoning takes place online, nevertheless, an inferentially-driven account of communication on this basis has to provide a model that explicates the concept of 'understanding' as effectively analysed through a logical system that implements these assumptions (see e.g. Allott, 2005). So, even though such a system can be based on heuristics that short-circuit complex chains of inference (Grice, 2001, 17), the logical structure of the derivation of an output has to be transparent if the implementation of that model is to be appropriately faithful (see e.g. Grice, 1981, 187 for the calculability of implicatures). Agents that are not capable of grasping this logical structure independently cannot be taken to be motivated by such computations, except as an idealisation pending a more explicit account. On the other hand, ignoring in principle the actual mechanisms that implement such a system as a competence/performance issue or an issue involving Marr's (Marr, 1982) computational vs the algorithmic and implementational levels (see e.g. Stone (2005); Stone (2004); Geurts (to appear):ch4) does not shield one from charges of psychological implausibility: if the same effects can be accounted for with standard psychological mechanisms without appeal to the complex model then, by Occam's razor, such an account would be preferable, especially if subtle divergent predictions can be uncovered (see e.g. Horton and Gerrig, 2005).

The controversial notion of 'intention' as a psychological state has been explicated in terms of hierarchical planning structures (Bratman, 1990), a view generally adopted in AI models of communication (Cohen et al., 1990). As the Gricean individualistic view of speaker's intention being the sole determinant of meaning underestimates the role of the hearer, dialogue models have turned to Bratman's account of *joint intentions* to model participant coordination. In this account, joint intentions arise through the composition of appropriately coordinated individual intentions and a network of mutual beliefs. In this respect, the notion of Gricean conversational "cooperation" features prominently in H. Clark's account of communication: dialogue involves (intentional) joint actions built on the coordination of individual actions based on shared beliefs (*common ground*) (see e.g. Clark, 1996). Hence, a strong Gricean element still underlies reasoning about speakers' intentions and meaning even though now supported by an account in terms of joint action and conversational structure. Thus, even here, the move from individualistic accounts of action, planning and intention to joint action and coordination in dialogue sees the latter as derivative and raises important philosophical and psychological issues that challenge received views of meaning and language.

1.2 Re-evaluating Intentionalism

The first type of challenge comes from views that have emerged under the influence of late Wittgensteinian ideas on language or the refutation of any rigid distinction between natural and non-natural meaning. Prominent amongst these are Millikan's teleosemantic approach to language content (Millikan, 2005) and Brandom's social-inferential account of communication (Brandom, 1994) which severally target aspects of Gricean and neo-Gricean conceptions of communication.

Millikan argues against Gricean approaches to communication from a naturalistic point of view. She argues that the standard Gricean view, with its heavy emphasis on mind-reading which is demonstrably not achievable by small children (see below sections 1.3, 1.4), turns the process of language acquisition, a heavily context-dependent process, into a mystery. Unlike the Gricean conception of meaning$_{NN}$ which rules out causal effects on the audience, e.g. involuntary responses in the hearer, Millikan's account, to the contrary, examines language and communication on the basis of phenomena studied by evolutionary biology, with linguistic understanding seen as analogous to direct perception rather than reasoning:[2] Objects of direct ordinary perception, e.g. vision, are not less abstract than linguistic meanings. Both require contextual filling in through processing of the incoming data in order to be comprehended; yet, in the case of ordinary perception, this processing obviously does not require considering someone's intention. An analogous assumption can then be made as regards linguistic understanding, so that the resolution of underspecified input in context would not require considering interlocutors' mental states as a necessary ingredient. Millikan then provides an account of linguistic meaning in a continuum with natural meaning based on the *function* that linguistic devices have been selected to perform (their survival value). These functions are defined through what linguistic entities are supposed to do (not what they normally do or are disposed to do) so that "function", in Millikan's sense, becomes a normative notion. Norms of language, "conventions", are uses that had survival value, and meaning is thus equated with function. In contrast then to Bratman's account of intentional action which sees the planning structures involved as distinctive of rational agents, distinguishing them from entities exhibiting merely purposive behaviour (see e.g. Bratman, 1999, 5), in Millikan's naturalistic perspective, function, i.e. meaning, does not depend upon speaker intentions. Nonetheless, speakers indeed can be conceived as behaving purposefully in producing tokens of linguistic devices (as hearts and kidneys behave purposefully) but without representing hearers' mental states or having intentions about hearers' mental states (see also Csibra and Gergely, 1998; Csibra, 2008). Similarly, hearers understand speech through direct perception of what the speech is about without necessary reflection on speaker intentions.

Of course, adults can, and often do, use reflections about the interlocutor's mental states; but this is not a necessary ingredient for meaningful interaction. Gricean mechanisms, that is, can be invoked but only as derivative or in cases of failure of the normal functioning of the primary mechanisms involved in the recovery of meaning, such as deception etc. From this perspective, what the Schiffer and Strawson scenarios show is that Gricean assumptions are on the wrong footing as a foundation for accounts of communication: generalising from these elaborate cases to cases of ordinary interaction is like taking hallucinations as the basis of an account of veridical perception (for a rejection of this view in the domain of perceptual experiences see e.g. McDowell, 1982). It is then no wonder that similar paradoxes are generated, e.g. the *mutual knowledge paradox* (Clark and Marshall, 2002) according to which interlocutors have to compute an infinite series of beliefs in finite time. The dilemma here is that there is plenty of evidence for *audience design* in language production, a type of cooperative behaviour, posing the problem of how to model the interlocutors' abilities allowing them to achieve this during online processing. But the solution to such problems ideally should not replicate that problematic structure (see e.g. Clark and Marshall (2002), who assume that interlocutors carry around detailed models of the people they know which they consult when they come to interact with them). Replacing such accounts with a psychological per-

2. The strict dichotomy between "meaning$_{NN}$" and "showing" has also been disputed within Relevance Theory (see e.g. Wharton, 2003).

spective that focuses on the mechanisms involved can undercut the intractability of such solutions by invoking independently established low-level memory mechanisms that provide explanation of how people appear to achieve "audience designed" productions without in fact constructing explicit models of the interlocutor (see e.g. Horton and Gerrig (2005) where retrieval of ordinary episodic memory traces predicts both conformity and deviation from the dictates of the common ground idealisation in experimental settings). Moreover, by taking seriously the linguistic resources available to the interlocutors, research in Conversational Analysis has revealed that when these low-level mechanisms fail there are dedicated socially-controlled devices for repairing coordination (see also Clark, 1996; Ginzburg, forthcoming), devices which allow for a form of externalised inference as regards the interlocutors' purposes.

An alternative account of communication combining Gricean and Millikanesque perspectives is that of Recanati (2004), which makes Gricean higher-order intention recognition a prerequisite only for implicature reconstruction. For what he terms "primary processes", on the other hand, Recanati adopts Millikan's account of understanding-as-direct-perception for the pragmatic processes that are involved in the determination of the truth-conditional content of an underspecified linguistic signal. These processes are blind and mechanical relying on 'accessibility' so that no inference or reflection of speaker's intentions and beliefs is required. It is only at a second stage, for the derivation of implicatures, that genuine reasoning about mental states comes into play.

Brandom (1994) also eschews the individualistic character of accounts of meaning espoused by the Gricean perspective, as part of his rationalist programme for semantics/pragmatics, and, more generally, philosophy. But unlike Millikan (and Recanati), Brandom analyses meaning/intentionality as arising out of linguistic social practices, with meaning, beliefs and intentions all accounted for in terms of the linguistic game of giving and asking for reasons, a view adopted in the domain of computational semantics by Kibble (2006). The guiding principle behind such social, non-intentionalist explanations of communication and dialogue understanding is to replace mentalist notions such as 'belief' with public, observable practical and propositional 'commitments', in order to resolve the problems arising for dialogue models associated with the intersubjectivity of beliefs and intentions, i.e. the fact that such private mental states are not directly observable and available to the interlocutors. A further motivation arises from the fact that it has been shown that beliefs, goals and intentions underdetermine what "rational" agents will do in conversation: social obligations or conversational rules may in fact either displace beliefs or intentions as the motivation for agents' behaviour or enter as an additional explanatory factor (Traum & Allen 1994). Brandom's account presents an inferentialist view of communication which seeks to replace mentalist notions such as belief with public, observable practical and propositional commitments. Under this view (as in Asher & Lascarides 2008), commitment does not imply 'belief' in the usual sense. A speaker may publicly commit to something which she does not believe. And 'intention' can be cashed out as the undertaking of a practical commitment or a reliable disposition to respond differentially to the acknowledging of certain commitments.[3]

From our point of view, the advantage of such non-individualistic, externalist accounts (see also Burge, 1986) is that, in not giving supremacy to an exclusively individualist conception of psychological processes, they break the presumed exhaustive dichotomy between behaviourist and mentalist accounts of meaning and behaviour (see e.g. Preston, 1994) or code vs. inferential models

3. An intermediate position is presented by Lascarides and Asher (2009); Asher and Lascarides (2008) who also appeal to a notion of *public commitment* associated with dialogue moves but which they link to a parallel cognitive modelling component based on inference about private mental states (see also Traum and Allen, 1994; Poesio and Traum, 1997).

of communication (see e.g. Krauss and Fussell, 1996). Instead, ascribing contents to behaviours is achieved by supra-individual social or environmental structures, e.g. conventions, "functions", practices, routinisations, that act as the context that guides agents' behaviour. The mode of explanation for such behaviours then does not enforce a representational component, accessible to individual agents, that analyses such behaviours in folk-psychological mentalistic terms, to be invoked as an explanatory factor in the production and interpretation of social action. Individual agents instead can be modelled as operating through low-level mechanistic processes without necessary rationalisation of their actions in terms of mental state ascriptions (see e.g. Barr (2004) for the establishment of conventions and Pickering and Garrod (2004) for coordination). This view is consonant with recent results in neuroscience indicating that notions like intentions, agency, voluntary action etc. can be taken as post hoc confabulations rather than causally efficacious (work by Benjamin Libet, John Bargh and Read Montague, for a survey see Wegner, 2002): according to these results, when a thought which occurs to an individual just prior to an action, is seen as consistent with that action, and no salient alternative causes of the action are accessible, the individual will experience conscious will and ascribe agency to themselves.

Accordingly, when examining human interaction, and more specifically dialogue, notions like intentions and beliefs may enter into common sense psychological explanations that the participants themselves can invoke and manipulate, especially when the interaction does not run smoothly. As such, they DO operate as resources that interlocutors can utilise explicitly to account for their own and others' behaviour. In this sense, such notions constitute part of the metalanguage participants employ to make sense of their actions in conscious, often externalised reflections (see e.g. Heritage (1984); Mills and Gregoromichelaki (2010); Healey (2008), section 2.2 below). Cognitive models that elevate such resources to causal factors in terms of plans, goals etc. either risk not doing justice to low-level mechanisms that implement the epiphenomenal effects they describe, or they frame their provided explanations as competence/computational level descriptions (see e.g. Stone, 2005, 2004). The stance such models take may be seen as innocuous preliminary idealisation, but this is acceptable only in the absence of either emerging internal inconsistency or alternative explanations that subsume the phenomena under more general assumptions. For example, there are well-known empirical/conceptual problems with the reduction of agent coordination in terms of Bratman's joint intentions (Searle, 1990; Gold and Sugden, 2007);[4] and there are also psychological/practical puzzles in cognitive/computational implementations in that the plan recognition problem is known to be intractable in domain-independent planning (Chapman, 1987). But, more pertinently for our concerns, cashing out communicative intentions in causal terms via the planning metaphor (Bratman, 1990) ignores the fact that the kinds of representations interlocutors actually employ to perform and interpret action do not explicitly deal with intentions or plans (unless these are explicit, conscious deliberations). As argued by Suchman (1987/2007); Agre and Chapman (1990), instead of taking plans and intentions as causal factors inside the agents' head guiding their action, they should be seen as arising as explicit articulations of antecedent conditions and consequences of past or future action that account for it in a way that can be made sense of by the agents themselves or the interpreters of their behaviours. In that respect, plans and intention attribution have a genuine explanatory role to play in human cognition and interaction but we see no reason to assign similar status, in addition, to mechanisms that are formulated in non-mentalistic, mechanistic terms to which agents have no conscious access. From this perspective, these mechanisms do not display

4. In addition such accounts of coordination are not general enough in that they are discontinuous with explanations of collective actions, in e.g. crowd coordination, individuals walking past each other on the sidewalk, etc.

identical functional roles with folk-psychological concepts (cf Stone 2004) and such metaphorical appropriation of such notions in fact obscures the actual function that explicit use of plan and intention attributions play in the agents' cognition.[5] These conclusions can be further substantiated on the basis of empirical and psychological evidence to which we now turn.

1.3 Re-evaluating intentionalism: empirical evidence

Buttressing these foundationalist arguments is a range of psycholinguistic research suggesting that recognition of intentions is an unduly strong psychological condition to impose as a prerequisite to effective communication. First, there is the problem of autism and related disorders. Autism, despite being reliably associated with inability (or at least markedly reduced capacity) to envisage other people's mental states, is not a syndrome precluding first-language learning in high-functioning individuals (Glüer and Pagin, 2003). Secondly, language acquisition across children is established well before the onset of ability to recognise higher-order intentions (Wellman et al., 2001), as evidenced by the so-called 'false-belief task' which necessitates the child distinguishing what they believe from what others believe (Perner, 1991). Given that language-learning takes place very largely through the medium of conversational dialogue, these results appear to show that at least communication with and by children cannot rely on higher-order intention recognition.

There is also very considerable independent evidence that even though adults are able to think about other people's perspectives, they are significantly influenced by their own point of view (*egocentrism*) (Keysar, 2007). This suggests that the complex hypotheses required by Gricean reasoning in communication may not reliably be constructed by adults either.[6] This is corroborated by an increasingly large body of research demonstrating that Gricean "common ground" is not a necessary building block in achieving coordinative communicative success: speakers regularly violate shared knowledge at first pass in the use of anaphoric and referential expressions which supposedly demonstrate the necessity of established common ground (Keysar, 2007, a.o.).[7] In accordance with these results, it is a regular occurrence in conversation that both speakers and hearers may elect not to make use of what is well established shared knowledge. On the one hand, in selecting an interpretation, a hearer may fail to check against consistency with what they believe the speaker could have intended (as in (1) where B construes the question in flagrant contradiction to what she knows A knows):

(1) A: Why don't you have bean chili?
B: Beef? You KNOW I'm a vegetarian [natural data]

Furthermore, the speaker's choice of anaphoric expression, supposedly restricted to well established shared knowledge, is regularly made in apparent neglect of what the hearer might take as salient:

5. In addition, it has been argued that use of such folk-psychological constructs are culture/occasion-specific (Du Bois, 1987; Duranti, 1988), hence should not be seen as underpinning general cognitive abilities.
6. Indeed, it is useful to note that even adults fail the false belief task, if it is a bit more complex (Birch and Bloom, 2007).
7. Though 'audience design' and coordination effects are regularly observed in experiments (see e.g. Hanna et al., 2003), these can be shown to result from general memory-retrieval mechanisms rather than as based on some common ground calculation based on metarepresentation or reasoning (see Horton and Gerrig, 2005; Pickering and Garrod, 2004).

(2) A having read out newspaper headline about Brown and Obama, upon reading next headline provides as follow-on:
A: *They*'ve received 10,000 emails.
B: Brown and Obama?
A: No, the Camerons. [natural data]

Given this type of example, checking in parsing or producing utterances that information is jointly held by the dialogue participants - the perceived *common ground* - cannot be a necessary condition on such activities.[8] One might want to characterise (1)-(2) as dysfunctional uses of language, impaired performance etc. But, firstly, there is psycholinguistic evidence that such neglect of common ground does not significantly impede successful communication and is not even detected by participants (Engelhardt et al., 2006, a.o.). Secondly, if indeed such data are set aside as unsuccessful acts of communication, one is left without an account of how people manage to understand what each other has said in these cases. But it is now well-documented that "miscommunication" not only provides vital insights as to how language and communication operate (Schegloff, 1979), but also facilitates dialogue coordination: as Healey (2008) shows, the local processes involved in the detection and resolution of misalignments during interaction lead to significantly more positive effects on measures of successful interactional outcomes (see also Brennan and Schober, 2001). In addition, these localised procedures lead to more gradual, group-level modifications, which in turn account for language change. Therefore, the Gricean and neo-Gricean focus on detecting speaker meaning as the sole criterion of communicative success misrepresents the goals of human interaction: miscommunication (which is an inevitable ingredient of interlocutors that do not share a priori common ground) and the specialised repair procedures made available by the structured linguistic and interactional resources available to interlocutors are the sole means that can guarantee intersubjectivity and coordination; and, as Saxton (1997) shows, in addition, such mechanisms, in the form of negative evidence and embedded repairs (see also Clark and Lappin, 2011), crucially mediate language acquisition (see also Goodwin, 1981, 170-171).

1.4 The weakening of Gricean assumptions

Such evidence has led to a move within Relevance Theory (RT) (Sperber and Wilson, 1995) weakening further the Gricean assumptions (Breheny, 2006). The relevance-theoretic view of communication is that the content of an utterance is established by a hearer relative to what the speaker could have intended (relative also to a concept of 'mutual manifestness' of background assumptions). This explanation involves meta-representation of other people's thoughts, but the process of understanding is effected by a mental module enabling hypothesis construction about speaker intentions. As noted by RT researchers, along with the communicated propositions, the context for interpretation falls under the speaker's communicative intention and the hearer selects it (in the form of a set of conceptual representations) on this basis. So, even though, unlike common ground, mutual manifestness of assumptions are in principle computable by conversational participants, and the interpretation process is not a rational one in the sense of Grice, it still remains the case that speaker meaning and intention are the guiding interpretive criteria which are implemented on mechanisms that have evolved to effect mind-reading. For this reason, Breheny argues that children in the initial

8. We are not claiming here that explanations for such phenomena cannot be given within standard Gricean models, especially since reasoning about speaker's intention is a form of non-demonstrative inference which can be expected to go wrong. The issue is whether such cases should be seen as deviant or not.

stages of language acquisition communicate relative to a weaker 'naive-optimism' strategy in which some context-established interpretation is simply presumed to match the speaker's intention, only coming to communicate in the full sense substantially later (see also Tomasello, 2008). In effect, this presents a non-unitary view of communication, which, based on the occasional sophistication that adult communicators exhibit, radically separates the abilities of adult communicators from those of children and high-functioning autistic adults.

On the other hand, given the known intractability of notions like planning recognition and common ground/mutual knowledge computation, computational models of dialogue, even when based on generally Clarkian theories of common ground, have also largely been developed without explicit high-order meta-representations of other parties' beliefs or intentions except where dealing with complex dialogue domains (e.g. non-cooperative negotiation, Traum et al., 2008). With algorithmically defined concepts such as *dialogue gameboard*, *QUD*, (Ginzburg, forthcoming; Larsson, 2002) and default rules incorporating rhetorical relations (Lascarides and Asher, 2009; Asher and Lascarides, 2008), the necessity for rational reconstruction of inferential intention recognition is largely sidestepped (though see Lascarides and Asher (2009); Asher and Lascarides (2008) for discussion). Even models that avow to implement Gricean notions (see e.g. Stone, 2005, 2004) have significantly weakened the Gricean reconstruction of the notion of "communicative intention" and meaning$_{NN}$ positing instead representations whose content does not directly reflect the logical structure (e.g. reflexive or iterative intentions) required by a genuine Gricean account. But, in our view, this is not the notion of rationality that Grice envisaged. And, as we said earlier (section 1.2), we see no reason to confuse the postulates of such models with the psychological constructs of the Gricean account. In fact, in many respects, these models are directly compatible with the view expressed here, namely, the need for low-level mechanistic explanations of joint action based on skills for collaboration and procedural knowledge.

2. Incrementality in Dialogue

Another set of major challenges to Gricean models of communication arise from the radical incrementality of processing in dialogue, and the incremental emergence of 'joint intentions' at the level of 'joint projects' (Bangerter and Clark, 2003).

2.1 Split utterances

The incrementality of on-line processing is now uncontroversial. It has been established for some considerable time now that language comprehension operates incrementally; and, standardly, psycholinguistic models assume that partial interpretations are built more or less on a word-by-word basis (see e.g. Sturt and Crocker, 1996). More recently, language production has also been argued to be incremental (Kempen and Hoenkamp, 1987; Levelt, 1989; Ferreira, 1996). Guhe (2007) further argues for the incremental conceptualisation of observed events resulting in the generation of preverbal messages in an incremental manner guiding semantic and syntactic formulation. In all the interleaving of planning, conceptual structuring of the message, syntactic structure generation and articulation, incremental models assume that information is processed as it becomes available, reflecting the introspective observation that the end of a sentence is not planned when one starts to utter its beginning (Guhe et al., 2000). In accordance with this, in dialogue, evidence for radical incrementality is provided not merely by the fact that participants incrementally "ground" each other's contribution (Allen et al., 2001) through *back-channel* contributions like *yeah*, *mhm*, etc. but

also by the observation that people clarify, repair and extend each other's utterances, even in the middle of an emergent clause:

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

In fact, such completions and continuations have been viewed by Herb Clark, among others, as some of the best evidence for cooperative behaviour in dialogue (Clark, 1996, 238). But even though, indeed, such joint productions demonstrate the communicators' skill to collaboratively participate in communicative exchanges, this ability to take on or hand over utterances raises the problem of the status of intention-recognition within human interaction. Firstly, on the Gricean assumption that pragmatic inference in dialogue operates on the basis of reasoning based on evidence of the interlocutor's intention, delivered by fixing the semantic propositional structure licensed by the grammar, the data in (3) cannot be easily explained, except as causing serious disruptions in normal processing. Secondly, on the assumption that communication necessarily involves recognising the propositional content intended by the speaker, there would be an expected cost for the original hearer in having to infer or guess this content before the original sentence is complete, and for the original speaker in having to modify their original intention, replacing it with that of another in order to understand what the new speaker is offering. But, wholly against this expectation, interlocutors very straightforwardly shift out of the parsing role and into the role of producer and vice versa as though they had been in that newly adopted role all along. Indeed, it is the case that such interruptions do sometimes occur when the respondent appears to have guessed what they think was intended by the original speaker, what have been called *collaborative completions*:

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

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

But this is not the only possibility: as (6)-(7) show, such completions by no means need to be what the original speaker actually had in mind:

(6) Morse: in any case the question was
Suspect: a VERY good question inspector [Morse, BBC radio 7]

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

In fact, such continuations can be completely the opposite of what the original speaker might have intended as in what we will call *hostile continuations* or *devious suggestions* which are nevertheless collaboratively constructed from a grammatical point of view:

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

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

Furthermore, as all of (4)-(10) show, speaker changes may occur at any point in an exchange (Purver et al., 2009), even very early, as illustrated by (10), with the clarification becoming absorbed into the final in-effect collaboratively derived content:

(10) A: They X-rayed me, and took a urine sample, took a blood sample. Er, the doctor
B: Chorlton?
A: Chorlton, mhmm, he examined me, erm, he, he said now they were on about a slide ⟨unclear⟩ on my heart. [BNC: KPY 1005-1008]

This phenomenon has consequences for accounts of both utterance understanding and utterance production. On the one hand, incremental comprehension cannot be based primarily on guessing speaker intentions: for instance, it is not obvious why in (6)-(9), the addressee has to have guessed the original speaker's (propositional) intention/plan before they offer their continuation.[9] On the other hand, speaker intentions need not be fully-formed before production: the assumption of fully-formed propositional intentions guiding production will predict that all the cases above where the continuation is not as expected, as in (6)-(9), would have to involve some kind of revision or back-tracking on the part of the original speaker. But this is not a necessary assumption: as long as the speaker is licensed to operate with partial structures, they can start an utterance without a fully formed intention/plan as to how it will develop (as the psycholinguistic models in any case suggest) relying on feedback from the hearer to shape their utterance (Goodwin, 1979). The importance of feedback in co-constructing meaning in communication has been already documented at the propositional level (the level of speech acts) within Conversational Analysis (CA) (see e.g. Schegloff, 2007). However, it seems here that the same processes can operate sub-propositionally, but only relatively to grammar models that allow the incremental, sub-sentential integration of cross-speaker productions. We turn to one such model next.

2.1.1 MODELLING THE INCREMENTALITY OF SPLIT UTTERANCES

The challenge of modelling the full word-by-word incrementality required in dialogue has recently been taken up, not merely within the Dynamic Syntax framework, a matter to which we return to

9. These are cases not addressed by DeVault et al. (2009), who otherwise offer a method for getting full interpretation as early as possible. Lascarides and Asher (2009); Asher and Lascarides (2008) also define a model of dialogue that partly sidesteps many of the issues raised in intention recognition. But, in adopting the essentially suprasentential remit of SDRT, their model does not address the step-by-step incrementality needed to model split-utterance phenomena.

in due course, but also by Poesio and Rieser (2010) (*P&R* henceforth). *P&R* set out a dialogue model for German, defining a thorough, fine-grained neo-Gricean model of dialogue interactivity that builds on an LTAG grammar base. Their primary aim is to model *collaborative completions*, as in (4) and (5), in cooperative task-oriented dialogues where take-over by the hearer relies on the remainder of the utterance taken to be understood or inferrable from mutual knowledge/common ground.Their account is an ambitious one in that it aims at modelling the generation and realisation of joint intentions which accounts for the production and comprehension of co-operative completions.

The P&R model hinges on two main areas: the assumption of recognition of interlocutors' intentions according to shared joint plans (Bratman, 1992), and the use of incremental grammatical processing based on LTAG. With respect to the latter, this account relies on the assumption of a string-based level of syntactic analysis, for it is this which provides the top-down, predictive element allowing the incremental integration of such continuations. This assumption, however, would seem to impede a more general analysis, since there are cases where split utterances cannot be seen as an extension by the second contributor to the proffered string of words/sentence:

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

(12) with smoke coming from the kitchen:
A: I'm afraid I burnt the kitchen ceiling
B: But have you
A: burned myself? Fortunately not.

In (11), the string of words that the completion yields is not at all what either participant takes themselves to have constructed, collaboratively or otherwise. And in (12) also, even though the grammar is responsible for the dependency that licenses the reflexive anaphor *myself*, the explanation for B's continuation in the fourth turn of (12) cannot be string-based as then *myself* would not be locally bound (its antecedent is *you*). Moreover, in LTAG, P& R's selected syntactic framework, words are defined in terms of syntactic/semantic pairings, relative to a given head, with adjuncts as a means of splitting these. Yet, as (7)-(12) indicate, utterance take-over can take place without a head having occurred prior to the split (see also Purver et al 2009, Howes et al this volume), and even across split dependencies (in (13) between an NPI and its triggering environment):

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

Given that such dependencies are defined grammar-internally, the grammar has to be able to license such split-participant realisations. But string-based grammars cannot account straightforwardly for many types of split utterances except by defining each part as sentential in its own right.

Furthermore, if the attempt is to reconstruct speaker's intentions as part of the interpretation recovered, as P&R explicitly advocate, there is the additional problem that such fragments can play multiple roles at the same time (in (5), (7), (11): question/completion/acknowledgment/answer). Not only that but the CA sequential structures (speech acts) normally taken to underpin coherence among propositional turn units, in fact, also operate within such collaborative constructions. For example, such completions might be explicitly invited by the speaker thus forming a question-answer pair:

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

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

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

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

(18) Anon 1: Yeah, the all-weather
Joyce: gallops?
Anon: yeah, wh what we call the all-weather gallop. [BNC, HDH: 169-171]

Within the P&R model, such multifunctionality would not be capturable except as a case of ambiguity or by positing hidden constituent reconstruction that has to be subject to some non-monotonic build-and-revise strategy that is able to apply even within the processing of an individual utterance. In addition, in fact, in some contexts, invited completions have been argued to exploit the vagueness of the speech act involved to avoid overt elicitation of information (Ferrara, 1992):

(19) (Lana = client; Ralph = therapist)
Ralph: Your sponsor before ...
Lana: was a woman

It has to be said that the P&R account is not intended to cover such data, as the setting for their analysis is one in which participants are assigned a collaborative task with a specific joint goal, so that joint intentionality is fixed in advance and hence anticipatory computation of interlocutor's intentions can be fully determined; but such fixed joint intentionality is decidedly non-normal in dialogue and leaves any uncertainty or nondeterminism in participants' intentions an open challenge. Nonetheless, by employing a dynamic view of the grammar, the P&R account marks a significant advance in the analysis of such phenomena.

2.1.2 FRAGMENTS AS INCOMPLETE SENTENCES?

Relative to any other grammatical framework, dialogue exchanges involving incremental split utterances of any type are even harder to model, given the near-universal commitment to a static performance-independent methodology. First of all, in almost all standard grammar frameworks, it is usually the sentence/proposition that is the unit of syntactic/semantic analysis. Fragments then are assigned sentential analyses with semantics provided through ellipsis resolution involving abstraction operations as in Dalrymple et al. (1991) (see e.g. Purver, 2004; Ginzburg and Cooper, 2004; Fernández, 2006). The abstraction is defined over a propositional content provided by the PREVIOUS context (in Ginzburg's terms the previously established *Question Under Discussion*) to yield appropriate functors to apply to the fragment. Of course, multiple options of appropriate

"antecedents" for elliptical fragments are usually available (one for each possible abstract). In consequence, some parsing mechanism that is defined to make reference to such a grammar-provided account of ellipsis must appeal to general pragmatic models having to do with recognizing the speaker's intention in order to select a single appropriate interpretation. But the intention recognition required for disambiguation is unavailable in sub-sentential split utterances in all but the most task-specific domains. This is because, in principle, attribution to any party of recognition of the speaker's intention to convey some specific propositional content is unavailable until the appropriate propositional formula is established. This is particularly clear where abstraction is required too early in the emergent proposition for there to be ANY appropriate abstract definable from context as that for which clarification is sought:

(10) A: They X-rayed me, and took a urine sample, took a blood sample. Er, the doctor
B: Chorlton?
A: Chorlton, mhmm, he examined me, erm, he, he said now they were on about a slide ⟨unclear⟩ on my heart. [BNC: KPY1005-1008]

Here, the only abstracts that are provided by the context in which B's clarificatory request of "Chorlton?" occurs are (informally) 'λ x. x took a blood sample', 'λy. y-x-rayed me', 'λ x. x took a urine sample', but none of these is the intended basis on which the fragment is uttered: what A presumes is that B can recover the interpretation as a request for clarification as to the identity of the doctor, and this is not of propositional type.[10] So such data constitute counter-examples for this style of account: at best, it remains incomplete, needing some other explanation for such early-placed clarifications.

Such collaboration without necessary recovery of Gricean intentions applies not only at the level of the sentence/turn but also at higher levels of discourse organisation ('joint projects') as we will discuss immediately below: people begin to interact in order to jointly achieve the completion of a cooperative task without having figured out what they are expected to do as would be predicted by planning models. Instead, they expect that, by engaging in the task, interactional routines will emerge that will guide their actions.

2.2 Emergent intentions: experimental evidence

While core pragmatic research has largely left on one side the phenomenon of collaborative construction of propositional information, the emergence of propositional contents in dialogue has been documented over many years in Conversation Analysis (CA) (see e.g. Lerner 2004). But both CA empirical research (Schegloff, 2007) and psycholinguistic experiments suggest that the same phenomenon can also be observed at higher levels of discourse organisation, the level of 'joint projects' (Bangerter and Clark, 2003). By probing the process of coordination in task-oriented dialogue experiments it can be demonstrated that notions of joint intentions and plans emerge gradually in a regular manner, rather than guiding utterance production and interpretation throughout.

Maze-game experiments (see Garrod and Anderson, 1987) provide a context in which, despite the high-level shared goal for the participants, the lower-level intentions/subplans over which they

10. It may be that Ginzburg and Cooper (2004)'s *constituent* clarification abstraction approach could be successfully employed at such early stages, as it requires only the presence of a recognised syntactic constituent rather than a full sentential proposition. But as currently formulated, it cannot apply to examples like (10) where the clarificational fragment *Chorlton* is lexically distinct from its antecedent *the doctor*.

have to coordinate are not given from the start. In experimental studies using the chat-tool methodology of Healey et al. (2003), this emergence of intentions during the course of the exchange was probed. This methodology allows controlled manipulations to be applied to the dialogue: in this case, artificial clarification requests were inserted by the server into the maze-game task which only one participant (A below) could see. A's response and the subsequent acknowledgment by the server were also not visible to the other participant:[11]

(20) A: I'm at the top (`Target turn`)
Server: top? (`Artificial turn by server`)
A: yes (`Response by A`)
Server: thanks (`Artificial ack. by server`)

In such exchanges, highly formulaic conventions emerge, revealing the high levels of coordination among participants. At late stages of a series of games, participants may have developed highly efficient elliptical exchanges like the following:

A: 4,5 2,6 1,4
B: 1,2 3,4 7,1
(21) A: 1,2
B: 4,5
A: 1,2 from Mills (2007)

The interpretation of these fragments crucially relies on the rich sequential structure of the joint project as even homonymous fragments like *1,2* above acquire distinct interpretations depending on their position in the sequence (e.g. the second *1,2* is interpreted as "I can get to 1,2" whereas the third as "I am now on 1,2", see Mills and Gregoromichelaki (2010)). Clarification requests surreptitiously inserted into the dialogue can then be interpreted in various ways as revealed by the participants' responses: sometimes receiving standard interpretations vis-a-vis content (Ginzburg and Cooper, 2004; Purver, 2004; Schlangen, 2004) as in (22); but sometimes as querying "what the turn is doing" in the sequence in which it occurs (Drew, 1997), as in (23):[12]

(22) A: 5, 6 (`Target turn`)
Server: what?/5? (`Artificial clarification`)
A: 5 along 6 across

(23) A: 5, 6 (`Target turn`)
Server: what?/5? (`Artificial clarification`)
A: go to 5 across, 6 is my switch

11. The chat tool is an experimental resource for carrying out investigation of dialogue, allowing fine-grained interventions over the communicative features of the interaction. Participants communicate through a familiar (instant-messenger-like) text-based interface. However, instead of passing turns directly to the appropriate chat clients, each turn is routed via a server. This information can then be used to trigger specific experimental interventions. For example, an artificial clarification request might be issued that appears to originate from another participant. The recipient responds to the clarification, and the server produces an acknowledgement, neither of which are seen by the other participant. Subsequent turns are then transmitted as normal. It has been shown that this can be done without disruption to the dialogue or detection by the participants.
12. For reasons of space, the following are amalgamations of real examples in the data.

The range of responses to such artificial requests were examined. This revealed a differential pattern in CR responses in early vs. late games. During the first few mazes, when the participants are relatively inexperienced in the task, CRs are interpreted as querying the referential import of the constituent concerned. At late stages of the interaction, however, both fragment and "what" CRs are interpreted significantly more frequently as concerning the intention or plan behind the target utterance, i.e. as questioning what the target-turn as a whole "is doing" in the sequence. These results can be interpreted as follows: Empirical CA analyses of the sequential coherence of conversation emphasise the importance of the turn-by-turn organisation of dialogue which allows juxtaposition of displays of participant understandings and provides structures for organised repair. Rather than interlocutors having to figure out each other's mental states and plans through metarepresentational means, conversational organisation provides the requisite structure for coordination. Similarly, as Garrod and Anderson (1987) observe, in maze-game experiments explicit negotiation is neither a preferential nor an effective means of coordination. If it occurs at all, it usually happens after participants have already developed some familiarity with the task. Hence, the Interactive Alignment model developed by Pickering and Garrod (2004) emphasizes the importance of tacit co-ordination and implicit common ground as the primary means of coordination. The establishment of routines and the significance of repair as externalised inference are also noted by Pickering and Garrod. The hypothesis that these implicit means, rather than intention recognition, were the primary method of coordination was further probed here by inserting artificial clarifications regarding intentions (*why?*) and observing the responses they receive at initial and later stages of a round of games (see Mills and Gregoromichelaki, 2008/in prep):

(24) `First few mazes`:

A: Go to the top right of the maze
Server: Why? (`artificial clarification`)
A: dunno/no idea

(25) A: Can you get to the top of the maze?
Server: Why? (`artificial clarification`)
A: Can you get to the top of the maze / Try it

Here too we observe that, at early stages, individuals display little recognition of specific intentions/plans underpinning their own utterance and explicit negotiation is either ignored or more likely to impede (see also Mills, 2007; Healey, 1997). This is because participants have not yet figured out the structure of the task, hence they do not have yet developed a metalanguage involving plan and intention attribution in order to explicitly negotiate their purposes. This implies that discursive constructs such as intentions need to emerge, even in such task-oriented joint projects. Initially, participants seem to follow trial-and-error strategies to figure out what the task involves. These strategies and the routines participants develop lead, at later stages of the maze-game, to highly co-ordinated, efficient interaction, as we saw earlier in (21), where, once task expertise is established, participants' utterances become highly contracted fragments (*telescoping*). As familiarity with the task and expertise increases, participants seem to disambiguate artificial clarification requests more and more as concerning "intentions" and plans:

(26) `Later mazes`:

A: 5, 6
Server: what?/5?/why? (`artificial clarification`)
A: because you've got to go there/you asked me to go there

These results appear to undermine both accounts of co-ordination that rely on an a priori notion of (joint) intentions and plans (see also Clark 1996) and also accounts which rely on some kind of strategic negotiation/agreement to mediate coordination. Instead, we take this as evidence that only at the late stages of a round of maze games can the presence of intentions and plans reliably guide the participants' interpretations and actions. However, even at these late stages, it is not necessary to assume that participants follow some explicit plan or have explicit intentions with respect to the interpretation of their fragmentary utterances.[13] The formulaic structure of their exchanges and the embodied responses that have developed, underpinned by the participants' increasing expertise with the task, makes it possible for them to still avoid having to work out each other's intentions to disambiguate the fragments, notwithstanding the potential to resolve any arising confusion by explicit appeal to intentions or plans. In any case, it would not be desirable to assume that the participants do not "communicate" at early stages of the interaction when they still have not figured out what the task involves and how it's structured. Hence, even in such task-specific situations, joint intentionality is not guaranteed *ab initio* but rather has to evolve incrementally with the increasing expertise.[14] These observations seem consonant with an alternative approach to planning and intention-recognition according to which forming and recognising such constructs is a subordinated activity to the more basic processes that underlie people's performance (see e.g. Suchman, 1987/2007; Agre and Chapman, 1990).

In accordance with this, in ordinary conversation, there is no guarantee that there is a genuinely shared plan, or that the way the shared utterance evolves is what either party had in mind to say at the outset, indeed obviously not, as otherwise exchanges like the ones in (4) and (7) etc would appear otiose. Instead, utterances are shaped genuinely incrementally and "opportunistically" according to feedback by the interlocutor (as already pointed out by Clark 1996). Grammatical integration of such joint contributions must therefore be flexible enough to allow such switches, with fragment resolutions occurring incrementally before computation of intentions is even possible.

3. Dynamic Syntax

In response to the challenge that such data provide, we turn to Dynamic Syntax (DS: Kempson et al 2001, Cann et al 2005) to consider whether forms of correlation between parsing and generation, as they take place in dialogue, can provide a basis for modelling recovery of interpretation in communicative exchanges without reliance on recognition of specific intentional contents. We set out a model of parsing and production mechanisms that makes it possible to show how, with speaker and hearer using incrementally the same mechanisms for construal, issues about interpretation choice and production decisions may be resolvable solely on the basis of feedback, without reflections on the other party's mental state. As we shall see, according to this account (Purver et al 2006), what underpins the smooth shift in all joint endeavours of conversation is the incremental,

13. Such fragments are argued to be assigned interpretations, through the routinisation of sets of actions, formulated as ad hoc idioms ('ad-hoc concepts', Carston (2002)) (see Mills and Gregoromichelaki, 2008/in prep).
14. Notably, the P&R data involve data collected after task training.

context-dependent processing shared by parsing and generation, and the tight coordination thereby achievable (similar assumptions underpin the model presented in Stone, 2004, 2005, even though distinct conclusions are drawn there as to its implications with respect to the issue of intention recognition in communication).

Instead of data such as (4)-(19) being problematic, extensive use of mechanisms across interlocutors illustrates the advantages of a DS-style incremental, dynamic account over static models. The incremental licensing of word processing modelled by DS, directly provides for the construction of restricted, contextually salient structural frames within which fragment construal/generation takes place. From a parsing perspective, this allows narrowing down of the threatening multiplicity of interpretations by incrementally weeding out possibilities en route to some commonly shared understanding. But the features of incrementality, predictivity/goal-directedness and context-dependent processing are built into the grammar architecture itself, rather than being external factors imposed by parsing/production mechanisms: each successive processing step relies on a grammatical apparatus which integrates lexical input with essential reference to the context in order to proceed. Under this low-level licensing of incrementally expanding strings and their interpretations, no mechanisms trigger high-level decisions about speaker/hearer intentions as part of the grammar itself. Rather, participants are modelled as gradually shaping propositional contents, on a word-by-word basis, drawing on subpersonal, synchronised mechanisms, without having to start with a fully-formed truth-evaluable content in mind. Such a view is buttressed by the fact that, as (7)-(12) show, neither party in such role-exchanges are able to know the eventual joint proposition in advance.

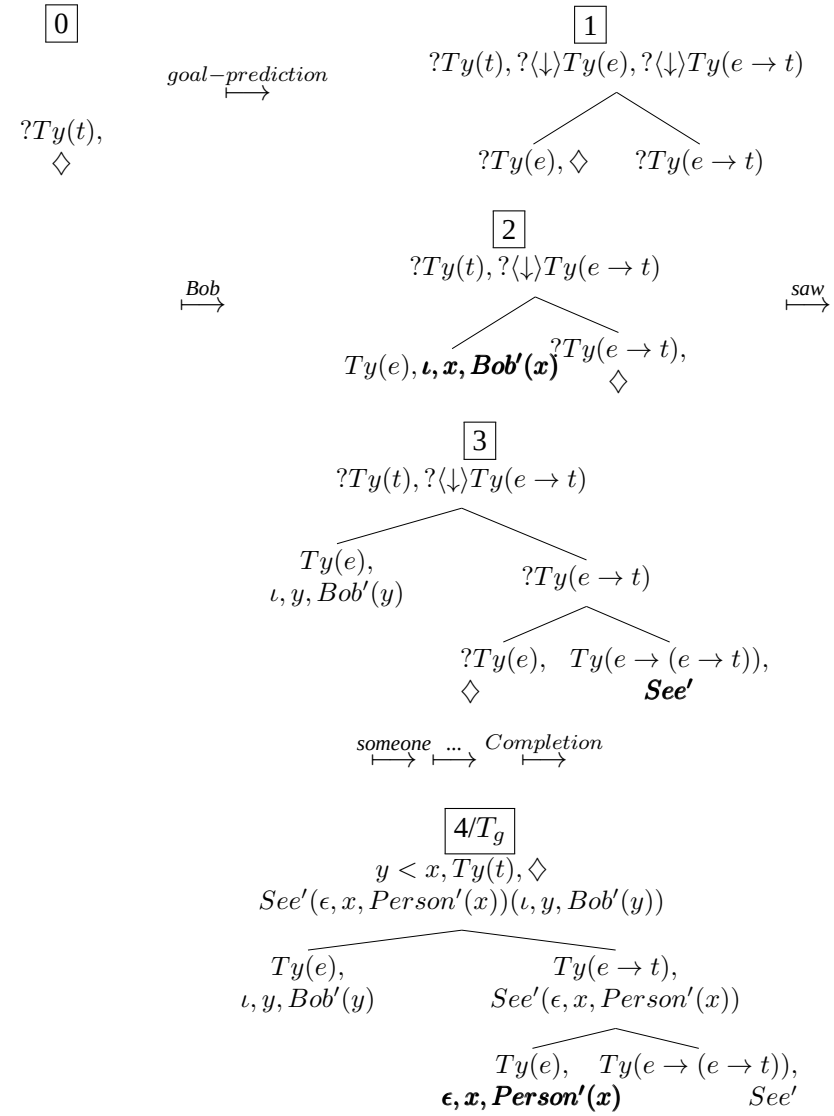
Our DS-based claim then is that communication involves taking risks without requiring mind-reading as an essential attribute: success in communication thus characteristically involves cycles of clarification/correction/extension/reformulation etc ("repair strategies") as essential subparts of the exchange. When modelled non-incrementally, such strategies might lead to the impression of non-monotonic repair and the need to revise some otherwise stable context. But pursued incrementally, within a goal-directed architecture, as we shall see, these do not constitute communication breakdown or disfluencies, but the normal mechanism of context construction, hypothesised update, and confirmation (see also Schegloff, 1979). By building on the assumption that successful communication may crucially involve subtasks of repair (Ginzburg, forthcoming), mechanisms for informational update that underpin interaction can be defined without reliance on (meta-)representing contents of the interlocutors' mental states as a precondition for successful communication. This is, emphatically, not to deny the rich human capacity for mind-reading but simply to argue that it is not a pre-requisite for effective communicative exchanges to take place.

3.1 Dynamic Syntax: the formalism

DS is a procedure-oriented framework modelling sequential processing. As is displayed in (28) by way of illustration, the build up of interpretation for (27) is monotonic and strictly word-by-word incremental:

(27) Bob saw someone

(28)



As (28) illustrates, the DS system provides mechanisms that enable the hearer to anticipate and therefore allow incremental word-by-word build-up of representations of content paired with some word string. Amongst such predictive steps are the construction by anticipation of a subject-predicate schema (stages 0-1 above, with requirements for a subject and predicate ($?\langle\downarrow\rangle Ty(e), ?\langle\downarrow\rangle Ty(e \rightarrow t)$) imposed as a very first step (not illustrated here), and their immediate construction at the second). Such a frame then makes possible the identification of the subject as some individual named Bob, via processing of the word *Bob* (stage 2), and then successive steps of identifying the predicate and its internal argument to be paired with verb and object noun-phrase respectively (stages 3-4). These updates then provide input to the compilation by labelled type-deduction of a propositional representation of content (stage 4). This then as a final step is subject to an algorithm

of evaluation determining how some assigned scope dependency choices are reflected in the constructed names (here the formula $S < x$ indicates that the existential term binding the variable x is taken as dependent on the event-term S (see Gregoromichelaki, 2011; Cann, forthcoming). The mechanisms for tree growth and evaluation are identically available to speakers, hence in generation. The only essential difference in production is that the modelling of a speaker's actions for tree-growth update involve a so-called "goal tree" (tree 4 in (28)) relative to which all intermediate construction steps have to be checked for commensurability, a checking step for which there may be no analogue in parsing (see section 3.2).

The notion of incrementality in DS is closely related to another of its features, the *goal-directedness/ predictivity* of BOTH parsing and generation (Demberg-Winterfors, 2010, see also). At each stage of processing, *structural predictions* are triggered that could fulfill the goals compatible with the input, in an underspecified manner. Representations of the conceptual structure of messages are given as binary trees, formally encoded with the tree logic *LOFT* Blackburn and Meyer-Viol (1994). LOFT is a modal logic with operators $\langle\uparrow\rangle, \langle\downarrow\rangle$ $\langle\uparrow_*\rangle, \langle\downarrow_*\rangle$ to define the relations of immediate and iterative domination, and to indicate node locations. What is novel about such trees is, on the one hand, that though they constitute a form of syntax, they are not inhabited by words of the language – they constitute structures inhabited by (lambda binding) formulae in the epsilon calculus, the selected semantic representation language. Furthermore, the mechanisms that define such progressive tree construction constitute the sole concept of natural-language syntax which the DS grammar provides. The system is goal-directed; and trees are constructed, by starting (in the context-independent case) from a radically underspecified goal, the *axiom* (the leftmost minimal tree in the illustration provided by (28), and proceeding through *monotonic updates* of partial or *structurally underspecified* trees until some tree is constructed from an input string in which all imposed goals and subgoals are met. Every node in a complete tree bears annotations that include the semantic formulae and their type information.

Crucial for expressing the goal-directedness are *requirements*, i.e. unrealized but expected node/tree specifications, indicated by '?' in front of annotations. As the axiom and its immediate subsequent update tree development in (28) indicate, requirements may also take a modal form, e.g. the constraint $?\langle\downarrow\rangle Ty(e \rightarrow t)$, which is a constraint that the daughter be a formula of predicate type. Requirements are essential to the tree-growth dynamics. All requirements must be satisfied if the construction process is to lead to a successful outcome, and, as indicated by the requirement for the predicate imposed at stage 2 in (28), these may not be satisfied until substantially later than the point at which they are imposed.[15]

Updates are carried out by means of applying both *computational* and *lexical actions*, which introduce and update nodes, and move the pointer. *Computational actions* govern general tree-constructional processes in a broadly top-down manner.[16] Lexical specifications, equally, induce actions that effect tree-development, providing annotations for nodes, in many cases also inducing the construction of further structure.[17] In the update from stage 2 to 3 in (28), for example, the set of lexical actions for the word *see* is applied, yielding the predicate subtree and its annotations. Sub-

15. The pointer, $\Diamond$, indicates the 'current node' in processing, namely the one to be processed next, a constraint which governs word order.
16. This is the characterisation of incrementality adopted by some psycholinguists under the appellation of *connectedness* (Sturt and Crocker (1996)): an encountered word always gets 'connected' to a larger, predicted, tree.
17. For cases of *dislocation*, DS employs *unfixed* nodes (not developed in this paper) which is indeed a core notion: such nodes are initially assigned structurally underspecified positions that are subsequently updated (at the point of the gap, in transformational parlance): see Kempson et al 2001, Cann et al 2005 among others.

sequent computational actions involve progressive labelled type-deduction decorating non-terminal nodes in the tree strictly bottom-up until the goal defined in the axiom is reached. Indeed all actions, computational and lexical, are defined in the same tree-growth vocabulary, so there is free intercalation of the various types of process. Thus *partial trees* grow incrementally, driven by procedures associated with particular words as they are encountered while conforming to top-down modal requirements on later development.

Central to the framework is the modelling of quantification as a process of term construction, using the *epsilon calculus* as the basic formula language (the epsilon calculus is the formal language that employs arbitrary-name terms in predicate logic natural-deduction proofs). All terms are of type e: *epsilon terms*, as illustrated by $(\epsilon, x, Consultant'(x))$. This term constitutes an arbitrary witness of the existentially quantified formula $\exists x.Consultant'(x)$, as defined by the following equivalence: $\psi(\epsilon, x, \psi(x)) \equiv \exists x.\psi(x)$

Notice how this equivalence yields the effect that an epsilon term invariably reflects its containing environment (the predicate ψ in the term's restrictor is a duplicate of the predicate applying on the term). The construction of such terms is induced by actions which incrementally, in part lexically, specify and collect up scope constraints of the form $x < y$ (to be understood as the term with variable y is dependent on the term with variable x). For example, indefinites project epsilon terms subject to the constraint that they are invariably dependent on either another quantifying expression or a term within the temporal specification; names, as iota terms (e.g. $\iota, x, Bob'(x)$), are, in contrast, taken to be epsilon terms of widest scope. A final algorithmic step yields the complex structure of the resulting terms as required in the equivalence: thus *A consultant arrived* is assigned a propositional formula

$$Arrived'(\epsilon, x, Consultant'(x))$$

is evaluated as

$$Consultant'(a) \wedge Arrived'(a)$$

where

$$a = (\epsilon, x, Consultant'(x) \wedge Arrived'(x))$$

The overall dynamics is thus one of growth in names as well as in structure.

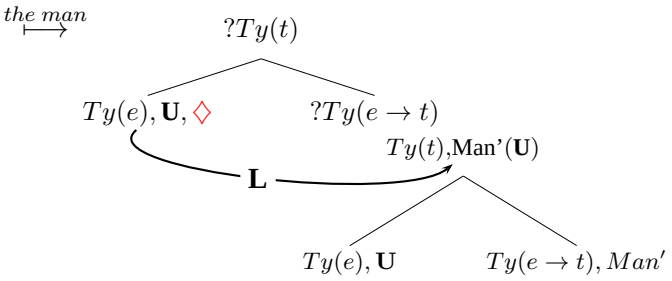
More radical underspecification of formulae at intermediate stages, equally associated with a process of growth, is lexically licensed, for example by pronouns, which act as simple place-holders for some possibly subsequent identification. These are defined as projecting a metavariable (notated as $\mathbf{U}, \mathbf{V}$ etc) as a place-holder for some value to be assigned, with an associated type specification, for pronouns $Ty(e)$. These invariably occur with an associated requirement for a fixed formula value (of the form $?\exists \mathbf{x} Fo(\mathbf{x})$), making such provision of a value essential to a successful outcome. Metavariables are substituted by other terms available in the context as part of the construction process, subject to locality restrictions differentiating e.g. pronouns and reflexives (for details see Cann et al., 2005; Kempson et al., 2001). A distinctive DS flavour lies in the license for a parse to proceed on the basis of such partial information. Indeed, given the type specification but lack of formula value in the processing of a pronoun, the value for such metavariables may be able to established somewhat later, as for example in expletive uses of pronouns, as in *It is likely that Geoff is wrong*.

In addition to the construction of individual predicate-argument structures, complex trees are obtained through a general tree-adjunction operation that licenses the construction of so-called LINKed trees. These are pairs of trees sharing information in the form of a shared term, each such tree a

subdomain in which labelled type-deduction takes place as in the simple structures. These provide a grammar-induced structural form of context. The construction processes determining and then updating such partial tree representations are used to model a range of phenomena.[18] For example, in taking definite NPs to be anaphoric, we define the definite article as introducing a metavariable as a partial term, inducing also construction of a LINK transition to allow the construction of a tree providing possibly complex information as a constraint ("presupposition") on the value to be substituted for the constructed partial term:

(29) The man smokes.

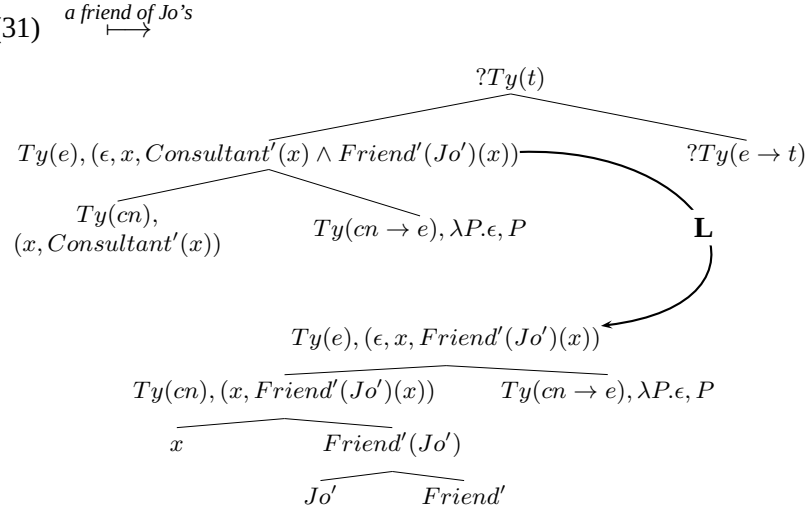
(30)



The structure on the subject node above is abbreviated as: $Ty(e), \mathbf{U}_{Man'(\mathbf{U})}$.

Appositional structures, as in *A consultant, a friend of Jo's, left*, can equally be established as inducing a pair of LINKed structures. A LINK transition is defined with the effect shown in (31) from a node of type e in which a preliminary epsilon term has been constructed onto a LINKed tree introduced with a requirement to develop a term using that very same variable:[19]

(31)



A twinned evaluation rule then combines the restrictors of two such paired terms to yield a composite term on the main tree (unlike the P&R account, this does not involve ambiguity of the head

18. The canonical case is relative clause construal (Kempson et al., 2001; Cann et al., 2005), where some type e term once processed becomes the context for the projection of one such LINKed structure, which, when completed, allows the pointer to return to that initial type e term, now enriched by the incorporation of information constructed upon such a LINKed (adjunct) tree.
19. In (31), we abbreviate the annotation $(\iota, y, Jo'(y))$ to Jo' for simplicity.

NP according to whether a second or subsequent NP follows). The fact that the first term has not been completed is no more than the term-analogue of the delaying tactic made available by expletive pronouns and extraposition-from-NP constructions, whereby a parse can proceed from some type specification of a node (with attendant metavariable as its formula value), but without completing (*evaluating*) that formula. Just as with expletives, this strategy allows term modification when the pointer returns from its sister node to that only partially constructed term immediately prior to compiling the decorations of its mother:

(32) A man has won, someone you know.

Such LINKed trees and their development set the scene for a general characterisation of context, ranging over possibly partial trees and their updates. *Context* in DS is defined as the storage of *parse states*, i.e., the storing of partial tree, word sequence parsed to date, plus the actions used in building up the partial tree. Formally, a parse state P is defined as a set of triples $\langle T, W, A\rangle$, where: T is a (possibly partial) tree; W is the associated sequence of words; A is the associated sequence of lexical and computational actions (Cann et al., 2007). At any point in the parsing process, the context $\mathcal{C}$ for a particular partial tree T in the set P can be taken to consist of: a set of triples $P' = \{\ldots, \langle T_i, W_i, A_i\rangle, \ldots\}$ resulting from the previous sentence(s); and the triple $\langle T, W, A\rangle$ itself, the subtree currently being processed. Anaphora and ellipsis construal generally involve re-use of formulae, structures, and actions from the set $\mathcal{C}$. All fragments illustrated above in (3)-(10) are processed by means of either extending the current tree, or by constructing LINKed structures with transfer of information among them so that one tree provides the context for another. Such fragments are licensed as wellformed by the grammar only relative to such contexts (Cann et al., 2007; Gargett et al., 2008; Kempson et al., 2009).

3.2 Parsing/generation coordination

This architecture allows a dialogue model in which generation and parsing function in parallel, following exactly the same procedure in the same order. Returning to (28), we now pick out the generation steps involved in producing *Bob saw Mary*, notated as (compressed) stages 0 to 4. As indicated earlier, generation of this utterance follows precisely the same actions and trees from left to right as in parsing, with the one additional filter, that the complete tree is available as a *goal tree* from the start (hence the labelling of the complete tree as T_g). The intuition this reflects is that the eventual message, in this simple context-independent case at least, is known in advance by the speaker and determines the choices to be made. What generation involves, in addition to parse steps, is reference to T_g to check whether each attempted generation stage (1, 2, 3, 4) is consistent with it. According to this algorithm, a *subsumption* check is carried out as to whether the current parse tree is monotonically extendible to T_g.[20] The trees 1-3 are licensed because, for each of these, the subsumption relation to T_g is maintained. Each time then the generator applies a lexical action, it is licensed to produce the word that carries that action only under successful subsumption check: at stage 3, for example, the generator processes the lexical action which results in the annotation See', and upon success and subsumption of T_g license to generate the word *see* ensues.

For processing split utterances, two more consequences are pertinent. First, there is nothing to prevent speakers initially having only a partial structure to convey, i.e. T_g may be a *partial* tree:

20. In fact, the goal tree for the speaker need only be subsumed by at least one parse step, and in all nonfinal steps in the generation process non-trivially.

this is unproblematic, as all that is required by the formalism is monotonicity of tree growth, and the subsumption check is equally well defined over partial trees. Second, the goal tree T_g may change during generation of an utterance, as long as this change involves monotonic extension; and continuations/reformulations/extensions across speakers are straightforwardly modelled in DS by appending a LINKed structure annotated with added material to be conveyed (preserving monotonicity) as in single speaker utterances:

(33) A friend is arriving, with my brother, maybe with a new partner.

Such a model under which the speaker and hearer essentially follow the same sets of actions, each incrementally updating their semantic representations, allows the hearer to mirror the same series of partial trees as the producer, albeit not knowing in advance the content of the unspecified nodes. Furthermore, not only can the same sets of actions be used for both processes, but also a large part of the parsing and generation algorithms is shared. In particular, the processing actions of both parsing and production involve the same progressive growth of partial tree representations, this being the only concept of "syntax" in the DS model. Even the concept of *goal tree*, T_g, may be shared between speaker and hearer, in so far as the hearer may have richer expectations relative to which the speaker's input is processed, as in the processing of a clarification question. Conversely, the speaker may have only a partial tree as T_g, relative to which they are seeking clarification.

In general, as no intervening level of syntactic structure over the string is ever computed, the parsing/generation tasks are more parsimonious in terms of representations than in other frameworks. Additionally, the top-down architecture in combination with partiality allows the framework to be (strategically) more radically incremental in terms of interleaving planning and production than is possible within other frameworks. On the one hand, there is one less level of representation to be computed, so no need for a complex step-by-step correlation of syntactic and semantic output, and no recourse either to some externally imposed parser to ensure such correlation. On the other hand, the licensing of partial structures allows articulation before a complete propositional goal has been determined and, therefore, interlocutor suggestions can be integrated without the need for revision.

4. Split utterances in Dynamic Syntax

Split utterances follow as an immediate consequence of these assumptions. For dialogues (7)-(12), A reaches a partial tree of what she has uttered through successive updates, while B as the hearer, follows the same updates to reach the same representation of what he has heard: they both apply the same tree-construction mechanism which is none other than their effectively shared grammar.[21] This provides B with the ability at any stage to become the speaker, interrupting to continue A's utterance, repair, ask for clarification, reformulate, or provide a correction, as and when necessary. According to DS assumptions, repeating or extending a constituent of A's utterance by B is licensed only if B, the hearer now turned speaker, entertains a message to be conveyed (a new T_g) that matches or extends in a monotonic fashion the parse tree of what he has heard. This message (tree) may of course be partial, as in (10), where B is adding a clarificational LINKed structure to a still-partially parsed antecedent, or it may complete the tree as in (12) and elsewhere.

Importantly, in DS, both A and B can now re-use the already constructed (partial) parse tree in their immediate context as a point from which to begin parsing and generation, rather than having

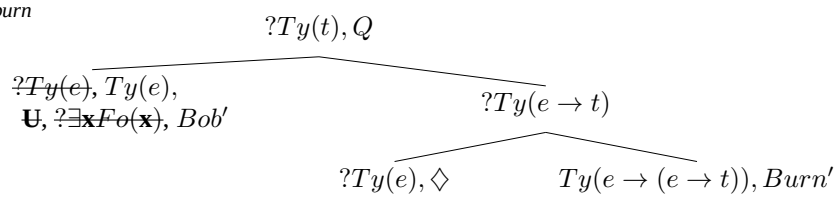
21. A completely identical grammar is, of course, an idealisation but one that is harmless for current purposes.

to rebuild an entirely novel tree or subtree. By way of illustration, we take a simplified variant of (12):

(34) Ann: Did you burn
Bob: myself? No.

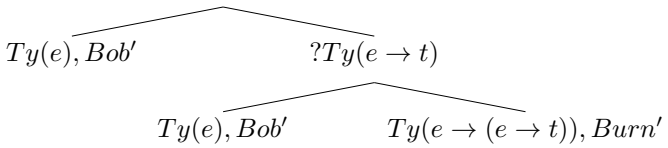
Here, of course, the reconstruction of the string as **Did you burn myself?* is unacceptable (at least with a reflexive reading of *myself*), illustrating the problem of purely syntactic accounts of split utterances. But under DS assumptions, with representations only of informational content, not of putative structure over strings of words, the switch of person is entirely straightforward. Consider the partial tree induced by parsing A's utterance *Did you burn* which involves a substitution of the metavariable projected by *you* with the name of the interlocutor/parser:[22]

(35) *Did you burn* $\longmapsto$



At this point, Bob can complete the utterance with the reflexive as what such an expression does, by definition, is copy a formula from a local co-argument node onto the current node, just in case that formula satisfies the conditions set by the person and number of the uttered reflexive, in this case, that it names the current speaker:

(36) *myself* $\longmapsto$



Hence the absence of a "syntactic" level of representation distinct from that of such semantic representations allows the direct successful integration of such fragments through the grammatical mechanisms themselves, rather than necessitating their analysis as sentential ellipsis.

Further, to illustrate how DS can sidestep the problems posed by abstraction accounts of ellipsis, we take a simplified version of (10):

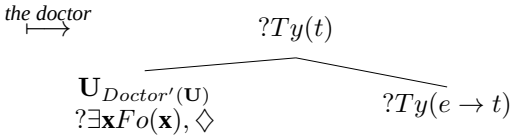
(37) A: The doctor

B: Chorlton?

After processing *the doctor* both A and B share a context comprising a partial tree as follows:

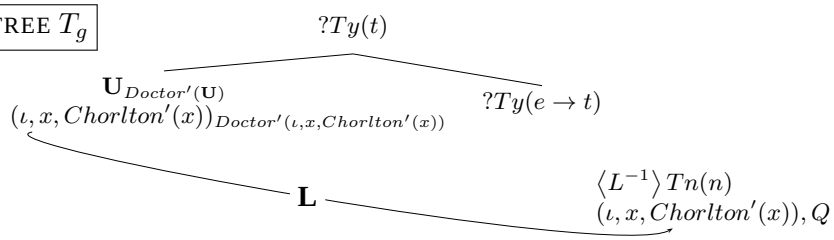
22. The feature Q on a decorated node is not taken to have a fixed speech-act content: given the range of acts achievable by interrogative structures (as diverse as *yes-no* questions, *wh* questions, tag questions, exclamatives, etc.) we take interrogative forms to encode a direction by the speaker to the hearer for a particular type of coordination, here notated simply as Q.

(38) A's/B's context:



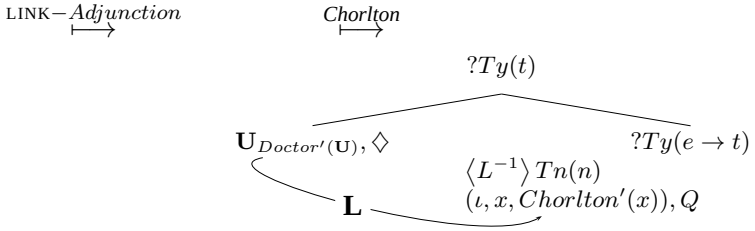
At the next stage of processing, let's assume that B fails to find a secure substitution for the metavariable **U** on the subject, thus being forced to request clarification if the requirement is to be satisfied. Notice that at this point A and B's contexts will diverge since A presumably knows who he's referring to, i.e. has a substituend for the metavariable introduced by the definite description. Now B's goal tree for his request for clarification is:

(39) B's GOAL-TREE T_g



The LINK transition, which accommodates an additional property (that the individual being talked about is named *Chorlton*), takes the partial tree in (38) as its context. In this context, with the pointer at the subject node, the building of a LINK relation is licensed and this is duly constructed. Now by uttering the word *Chorlton?* a new tree can be constructed for B which indeed subsumes the goal tree of (39):

(40) B's CONSTRUCTION-TREE



Now regular anaphoric substitution allows the metavariable **U** to be instantiated by the term ($\iota, x,$ $Chorlton'(x)$), indeed essentially, as otherwise the two nodes will not be developed as involving any shared term. The result of this process will be exactly the tree in (39) and speaker and hearer context trees will be identical at this point. As illustrated here, the most recent (partial) parse tree constitutes the most immediately available local "antecedent" for fragment resolution; hence no separate computation or definition of *salience* or speaker intention by the hearer is necessary for such incremental fragment construal. As in P&R, the mechanism is exactly that of apposition, the building of a LINKed structure, in this case, the result of that transition in its turn being used to provide a value for the metavariable place-holder associated with the definite article.

As we saw, the hearer, B, may respond to what he has constructed during interpretation, anticipating A's verbal completion as in (4) and (5). This is facilitated by the general predictivity/goal-directedness of the DS architecture since the parser is always predicting top-down goals (*requirements*) to be achieved in the next steps (see stage 2 of (28) or e.g. (38)). Such goals are indeed what

drives the search of the lexicon (*lexical access*) in generation, so a hearer who shifts to successful lexicon search before processing the anticipated lexical input provided by the speaker can become the generator and take over. In all the cases of split utterances, the original hearer is, indeed, using such anticipation to take over and offer a completion that, even though grammatically licensed, i.e. fitting the predicted structure of the context tree, might not necessarily be identical to the one the original speaker would have accessed had they been allowed to continue their utterance as in (7)-(9).[23] From this point of view, since both speakers and hearers are licensed to operate with partial structures, speakers can start an utterance without a fully-formed intention/plan as to how it will develop (as the psycholinguistic models in any case suggest) relying on feedback from the hearer to shape their utterance:

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

Hence the assumption of underspecified partial speaker contents before the beginning of articulation allows genuine collaboration in the construction of utterances (Goodwin, 1979), without necessarily having to resort to revision and backtracking.

5. Summary Evaluation

With grammar mechanisms defined as inducing growth of information that is used symmetrically and incrementally in both parsing and generation, the availability of derivations for genuine dialogue phenomena, like split utterances, from within the grammar, shows how core dialogue activities can take place without any other-party meta-representation at all (though use of reasoning over mental states is not precluded either). On this view, as we emphasised earlier, communication is not definitionally the full-blooded intention-recognising activity presumed by Gricean and post-Gricean accounts. Rather, speakers can, on this view, air propositional and other structures with no more than the vaguest of planning and commitments as to what they are going to say, expecting feedback to fully ground the significance of their utterance, to fully specify their intentions (see e.g. Wittgenstein, 1953, 337). Hearers, similarly, may signally fail to reconstruct putative intentions of their interlocutor as a filter on how to interpret the provided signal; instead, they are expected to provide evidence of how THEY perceive the utterance in order to arrive at a joint interpretation.

This view of dialogue, though not uncontentious, is one that has been extensively argued for, under distinct assumptions, in the CA literature. According to the proposed DS model of this phenomenon, the core ingredient of dialogue is incremental, context-dependent processing, implemented by a grammar architecture that reconstructs "syntax" as a goal-directed activity, able to seamlessly integrate with the joint activities people engage in. Incrementality is a facilitator both for allowing entirely individualistic decisions as to what to say and how, and also for, nevertheless, making possible a joint activity in which an emergent structure can unfold through the

23. It might be argued that back-channels (*Mhm*, *yes* etc), are problematic for this account. However, arguably, such signals do not encode recognition of intentional content even though this is often the interpretation assigned in context: even the canonical content-agreement device, *yes*, is systematically used to signal merely shared attention or license for the interlocutor to continue without an explicit propositional content necessarily being available:
(i) Sue (opening conversation): Tom/ (or knocking on Tom's door)
Tom (in response): Yes?
Sue: Are you busy?

requesting and receiving of feedback. The two properties that determine how such a weak underpinning can nevertheless yield the coordinative effect achieved in dialogue exchanges are the intrinsic predictivity/goal-directedness in the formulation of DS, and the fact that both parsing and production can have arbitrary partial goals, so that, in effect, both interlocutors are able to be building structures in tandem.

In particular, because of the assumed partiality of goal trees (messages), speakers do not have to be modelled as having fully-formed messages, reflecting intentions with propositional content, at the beginning of the generation task, but can instead be viewed as relying on feedback to shape their unfolding utterance. As goal trees are expanded incrementally, completions/repairs/feedback by the other party can be monotonically accommodated, even though they might not represent what the speaker would have uttered if not interrupted. As long as what emerges as the eventual joint content is some compatible extension of the original speaker's goal tree, it may be accepted as sufficient for the purposes to hand. Thus, in such an incremental model, repair procedures do not deviate from the normal processing mechanisms. In fact, the very possibility of some types of repair, e.g. mid-utterance self-repairs, requires the licensing of partial strings by the grammar (see e.g. Ginzburg et al., 2007). But further than that, an incremental syntactic model licenses strings and their interpretations on a word-by-word basis and thus can naturally integrate any "repairs" via the already assumed progressive accumulation of information. Hence "repair" phenomena naturally emerge as "coordination devices" (Clark, 1996), devices exploiting mutually salient contexts for achieving coordination enhancement. And jointly constructed content that is established through cycles of "miscommunication" and "repair" is more securely coordinated (see e.g. Healey, 2008, and section 1.3 above) and thus can form the basis of what each party considers shared cognitive context. In addition, the CA notion of 'sequentiality' that, in our view, can operate in dialogue both sub-propositionally (see (14)-(18)) and across turns through the turn-taking system has also the grammar as its most significant determinant (De Ruiter et al., 2006). The DS model captures this naturally as the notion of 'projection' (Schegloff, 1987) that underlies the possibility of harmonious turn-taking is integrated in the goal-directed/predictive architecture of the grammar, which requires the parser to constantly make assumptions as to what is licensed to follow, given an already established goal. Overall then, given that such core CA notions, generally assumed to reflect the efficiency, social aptness and highly organised nature of conversation, can be modelled as consequences of the operation of a low-level mechanism like the grammar, the view of communication that emerges here does not require essential grounding in having to recognize speaker's intentions, hence can be taken to be displayed equally by both young children and adults.

One might argue against this view of communication that the phenomenon of conversational implicature, in which speakers may direct hearers to the construction of additional hypotheses to yield indirect inferential effects, necessitates an essentially meta-representational view of communication and explicit representation of a speaker's intentions with respect to their interlocutor. However, there are alternative accounts of implicatures where, even though situated inference is involved, the explanations do not necessarily invoke an interlocutor metarepresentational component (see e.g. Gauker, 2001) (see also Arundale, 2008; Haugh, 2008). Such inferences might necessitate overt modelling of the interlocutor but not essentially. Accordingly, there is no restriction in the view proposed here on the types of representation participants may construct, so nothing precludes the construction of richer contexts to yield such effects.[24]

24. Such richer contexts and consequent derived implications could be modelled via the construction of appropriately term-sharing LINKed trees, whose mechanisms for construction are independently available in DS.

This then enables a new perspective on the relation between linguistic ability and the use of language, constituting a position intermediate between the philosophical stances of Millikan and Brandom, and one which is notably close to that of Recanati (2004). Linguistic ability is grounded in the control of (low-level) mechanisms (see e.g. Böckler et al., 2010) which enable the progressive construction of structured representations to pair with the overt signals of the language, used in conjunction with some generally available cognitive filter for determining particular choices made. The content of these representations is ascribed, negotiated and accounted for in context, via the interaction among interlocutors. Constructing representations of the other participants' mental states, though a possible means of securing communication, is by no means necessary. With Dynamic Syntax being a grammar formalism, we have not here had anything formal to say about the choice mechanism that selects interpretations from those made available through linguistic processing, although, given the Millikan view of communication and the psycholinguistic evidence favouring low-level mechanisms (e.g. Pickering and Garrod 2004, Horton and Gerrig 2005, Keysar 2007), we believe, along with Recanati, that such a mechanism does not operate through the implementation of Gricean assumptions. But, on this view, whatever the underpinnings of such a mechanism (e.g. *relevance* as defined in Sperber and Wilson (1995); or rhetorical relations or the cognitive logic of Lascarides and Asher (2009); Asher and Lascarides (2008)), it interacts stepwise with the implementation of the resources for interaction that are provided by the grammar (see also Ginzburg, forthcoming; Cooper and Ranta, 2008). Hence we suggest, contra Tomasello (2008), that we need to be exploring accounts of human communication as an activity involving emergent agent coordination without high-level mind-reading as a prerequisite skill.

Acknowledgments

We gratefully acknowledge helpful discussions with Alex Davies, Arash Eshghi, Chris Howes, Graham White, Hannes Rieser and Robin Cooper. We would also like to thank Hannes Rieser for editorial assistance and the reviewers, especially one of them, for detailed and incisive comments that vastly improved our clarity as regards the issues involved. This work is supported by The Dynamics of Conversational Dialogue (DynDial) ESRC-RES-062-23-0962; Leverhulme Trust Major Research Fellowship F00158BF for R. Cann; Marie Curie IOF fellowship (2010-2012) Grant no: PIOF-GA-2009-236632-ERIS for Gregory Mills.

References

P.E. Agre and D. Chapman. What are plans for? *Robotics and Autonomous Systems*, 6(1-2):17–34, 1990.

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

N. Allott. Paul Grice, reasoning and pragmatics. *UCL Working Papers in Linguistics*, pages 217–243, 2005.

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

N. Asher and A. Lascarides. Commitments, beliefs and intentions in dialogue. *Proceedings of Londial*, pages 35–42, 2008.

K. Bach. The semantics-pragmatics distinction: What it is and why it matters. *Linguistische Berichte*, 8(1997):33–50, 1997.

K. Bach and R.M. Harnish. *Linguistic communication and speech acts*. MIT Press, 1982.

A. Bangerter and H.H. Clark. Navigating joint projects with dialogue. *Cognitive Science*, 27(2): 195–225, 2003.

D.J. Barr. Establishing conventional communication systems: Is common knowledge necessary? *Cognitive Science*, 28(6):937–962, 2004.

SA Birch and P. Bloom. The curse of knowledge in reasoning about false beliefs. *Psychological Science*, 18(5):382, 2007.

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

A. Böckler, G. Knoblich, and N. Sebanz. Socializing Cognition. *Towards a Theory of Thinking*, pages 233–250, 2010.

R.B. Brandom. *Making it explicit: Reasoning, representing, and discursive commitment*. Harvard Univ Pr, 1994.

M. Bratman. *Faces of intention: Selected essays on intention and agency*. Cambridge Univ Pr, 1999.

Michael E. Bratman. What is intention? In P. Cohen, J. Morgan, and M. Pollack, editors, *Intentions in Communication*. MIT Press, 1990.

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

Richard Breheny. Communication and folk psychology. *Mind & Language*, 21(1):74–107, 2006.

S.E. Brennan and M.F. Schober. How Listeners Compensate for Disfluencies in Spontaneous Speech. *Journal of Memory and Language*, 44(2):274–296, 2001.

T. Burge. Individualism and psychology. *The Philosophical Review*, 95(1):3–45, 1986.

Ronnie Cann. Towards an account of the english auxiliary system. In R. Kempson, Gregoromichelaki E., and C. Howes, editors, *The Dynamics of Lexical Interfaces*, chapter 9. CSLI, forthcoming.

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

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

Robyn Carston. *Thoughts and Utterances: The Pragmatics of Explicit Communication.* Blackwell, 2002.

D. Chapman. Planning for conjunctive goals. *Artificial intelligence*, 32(3):333–377, 1987.

A. Clark and S. Lappin. *Linguistic Nativism and the Poverty of the Stimulus.* Wiley-Blackwell, 2011.

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

H.H. Clark and C.R. Marshall. Definite reference and mutual knowledge. *Psycholinguistics: Critical Concepts in Psychology*, page 414, 2002.

P.R. Cohen, J.L. Morgan, and M.E. Pollack. *Intentions in communication*. The MIT Press, 1990.

R. Cooper and A. Ranta. *Natural languages as collections of resources*. College Publications, 2008.

G. Csibra. Goal attribution to inanimate agents by 6.5-month-old infants. *Cognition*, 107(2):705–717, 2008.

G. Csibra and G. Gergely. The teleological origins of mentalistic action explanations: A developmental hypothesis. *Developmental Science*, 1(2):255–259, 1998.

M. Dalrymple, S. M. Shieber, and F. C. N. Pereira. Ellipsis and higher-order unification. *Linguistics and Philosophy*, 14(4):399–452, 1991.

J.P. De Ruiter, H. Mitterer, and N.J. Enfield. Projecting the end of a speakers turn: A cognitive cornerstone of conversation. *Language*, 82(3):515–535, 2006.

V. Demberg-Winterfors. *A broad-coverage model of prediction in human sentence processing*. PhD thesis, 2010.

D. DeVault, K. Sagae, and D. Traum. Can I finish? In *Proceedings of the SIGDIAL 2009 Conference*, pages 11–20, London, UK, 2009.

P. Drew. Open'class repair initiators in response to sequential sources of troubles in conversation. *Journal of Pragmatics*, 28(1):69–101, 1997.

J.W. Du Bois. Meaning Without Intention: Lessons from Divination. *Papers in Pragmatics*, 1(2), 1987.

A. Duranti. Intentions, language, and social action in a Samoan context. *Journal of Pragmatics*, 12 (1):13–33, 1988.

P.E. Engelhardt, K.G.D. Bailey, and F. Ferreira. Do speakers and listeners observe the Gricean Maxim of Quantity? *Journal of Memory and Language*, 54(4):554–573, 2006.

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

K. Ferrara. The interactive achievement of a sentence: Joint productions in therapeutic discourse. *Discourse Processes*, 15(2):207–228, 1992.

V. Ferreira. Is it better to give than to donate: syntactic flexibility in language production. *Journal of Memory and Language*, 35:724–755, 1996.

A. Gargett, E. Gregoromichelaki, C. Howes, and Y. Sato. Dialogue-grammar correspondence in dynamic syntax. In *Proceedings of the 12th SEMDIAL (LONDIAL)*, 2008.

S. Garrod and A. Anderson. Saying what you mean in dialogue: A study in conceptual and semantic co-ordination. *Cognition*, 27:181–218, 1987.

C. Gauker. Situated inference versus conversational implicature. *Noûs*, 35(2):163–189, 2001.

B. Geurts. *Quantity implicatures*. Cambridge University Press, to appear.

J. Ginzburg. *The Interactive Stance: Meaning for Conversation*. forthcoming.

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

J. Ginzburg, R. Fernández, and D. Schlangen. Unifying self-and other-repair. In *Proceeding of DECALOG, the 11th International Workshop on the Semantics and Pragmatics of Dialogue (SemDial07)*, 2007.

K. Glüer and P. Pagin. Meaning theory and autistic speakers. *Mind & Language*, 18(1):23–51, 2003.

N. Gold and R. Sugden. Collective intentions and team agency. *Journal of Philosophy*, 104(3): 109–37, 2007.

C. Goodwin. The interactive construction of a sentence in natural conversation. *Everyday language: Studies in ethnomethodology*, pages 97–121, 1979.

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

E. Gregoromichelaki. Conditionals in dynamic syntax. In R. Kempson, Gregoromichelaki E., and C. Howes, editors, *The Dynamics of Lexical Interfaces*, chapter 8. CSLI, 2011.

P. Grice. Logic and conversation. *1975*, pages 41–58, 1975.

P. Grice. Presupposition and Implicature. *Radical Pragmatics*, pages 183–199, 1981.

P. Grice. *Aspects of Reason*. Oxford: Clarendon Press, ed. by r. warner edition, 2001.

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

M. Guhe, C. Habel, and H. Tappe. Incremental event conceptualization and natural language generation in monitoring environments. In *Proceedings of the first international conference on Natural language generation-Volume 14*, pages 85–92. Association for Computational Linguistics, 2000.

J.E. Hanna, M.K. Tanenhaus, and J.C. Trueswell. The effects of common ground and perspective on domains of referential interpretation. *Journal of Memory and Language*, 49(1):43–61, 2003.

M. Haugh. Intention in pragmatics. *Intercultural Pragmatics*, 5(2):99–110, 2008.

Patrick Healey. Expertise or expert-ese: The emergence of task-oriented sub-languages. In M.D. Shafto and P. Langley, editors, *Proceedings of the 19th Annual Conference of the Cognitive Science Society*, pages 301–306, Stanford, California, August 1997. Stanford University.

Patrick Healey. Interactive misalignment: The role of repair in the development of group sublanguages. In R. Cooper and R. Kempson, editors, *Language in Flux*. College Publications, 2008.

Patrick Healey, Matthew Purver, James King, Jonathan Ginzburg, and Gregory J. Mills. Experimenting with clarification in dialogue. In *Proceedings of the 25th Annual Meeting of the Cognitive Science Society*, Boston, Massachusetts, August 2003.

J. Heritage. *Garfinkel and ethnomethodology*. Polity, 1984.

W.S. Horton and R.J. Gerrig. Conversational common ground and memory processes in language production. *Discourse Processes*, 40(1):1–35, 2005.

G. Kempen and E. Hoenkamp. An incremental procedural grammar for sentence formulation. *Cognitive Science*, 11(2):201–258, 1987.

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

Ruth Kempson, Eleni Gregoromichelaki, and Yo Sato. Incrementality, speaker/hearer switching and the disambiguation challenge. In *Proceedings of European Association of Computational Linguistics proceedings.*, 2009.

B. Keysar. Communication and miscommunication: The role of egocentric processes. *Intercultural Pragmatics*, 4(1):71–84, 2007.

R. Kibble. Reasoning about propositional commitments in dialogue. *Research on Language & Computation*, 4(2):179–202, 2006.

R. M. Krauss and S. R. Fussell. Social and psychological models of interpersonal communication. In E.T. Higgins and A.W. Kruglanski, editors, *Social Psychlogy: Handbook of Basic Principles*, pages 655–701. Guildford, 1996.

S. Larsson. *Issue-based Dialogue Management*. PhD thesis, Göteborg University, 2002. Also published as Gothenburg Monographs in Linguistics 21.

A. Lascarides and N. Asher. Agreement, disputes and commitments in dialogue. *Journal of Semantics*, 26(2):109, 2009.

G.H. Lerner. On the syntax of sentences-in-progress. *Language in Society*, pages 441–458, 1991.

G.H. Lerner. Collaborative turn sequences. In *Conversation analysis: Studies from the first generation*, pages 225–256. John Benjamins, 2004.

W.J.M. Levelt. *Speaking: From intention to articulation*. Mit Pr, 1989.

S. C. Levinson. *Pragmatics*. Cambridge Textbooks in Linguistics. CUP, 1983.

D. Marr. *Vision: A computational investigation into the human representation and processing of visual information*. Henry Holt and Co., Inc. New York, NY, USA, 1982.

J. McDowell. Criteria, defeasibility, and knowledge. In *Proceedings of the British Academy*, volume 68, pages 455–79, 1982.

R.G. Millikan. *Language: a biological model*. Oxford University Press, USA, 2005.

G. J. Mills. *Semantic co-ordination in dialogue: the role of direct interaction.* PhD thesis, Queen Mary University of London, 2007.

G. J. Mills and E. Gregoromichelaki. Coordinating on joint projects. based on talk given at the Coordination of Agents Workshop, Nov 2008, KCL, 2008/in prep.

G. J. Mills and E. Gregoromichelaki. Establishing coherence in dialogue: sequentiality, intentions and negotiation. In *Proceedings of the 14th SemDial, Pozdial*, 2010.

J. Perner. *Understanding the representational mind*. MIT Press Cambridge, MA, 1991.

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

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

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

B. Preston. Behaviorism and mentalism: Is there a third alternative? *Synthese*, 100(2):167–196, 1994.

M. Purver. *The Theory and Use of Clarification Requests in Dialogue*. PhD thesis, University of London, 2004.

M. Purver, Chris Howes, Eleni Gregoromichelaki, and Patrick Healey. Split utterances in dialogue: a corpus study. 2009.

F. Recanati. *Literal meaning*. Cambridge Univ Pr, 2004.

M. Saxton. The contrast theory of negative input. *Journal of Child Language*, 24(01):139–161, 1997.

E.A. Schegloff. The relevance of repair to syntax-for-conversation. *Syntax and semantics*, 12: 261–286, 1979.

E.A. Schegloff. Recycled turn beginnings. *Talk and social organization*, pages 70–85, 1987.

E.A. Schegloff. *Sequence organization in interaction: A primer in conversation analysis I*. Cambridge Univ Pr, 2007.

S.R. Schiffer. *Meaning*. Oxford University Press, USA, 1972.

D. Schlangen. Causes and strategies for requesting clarification in dialogue. In *Proceedings of the 5th SIGdial Workshop on Discourse and Dialogue*, Boston, Massachusetts, April 2004. Association for Computational Linguistics.

J. R. Searle. Collective intentions and actions. In Philip R. Cohen, J. Morgan, and M. E. Pollack, editors, *Intentions in communication*, pages 401–415. MIT Press, 1990.

D. Sperber and D. Wilson. *Relevance: Communication and Cognition (2nd editn)*. Blackwell, Oxford, 1995.

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

M. Stone. Communicative intentions and conversational processes in human-human and human-computer dialogue. In J. Trueswell and M. Tanenhaus, editors, *Approaches to Studying World-Situated Language Use*, pages 39–70. MIT Press, 2005.

PF Strawson. Intention and convention in speech acts. *The Philosophical Review*, 73(4):439–460, 1964.

P. Sturt and M. Crocker. Monotonic syntactic processing: a cross-linguistic study of attachment and reanalysis. *Language and Cognitive Processes*, 11:448–494, 1996.

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

M. Tomasello. *Origins of Human Communication*. MIT, 2008.

D. Traum and J. Allen. Towards a formal theory of repair in plan execution and plan recognition. *Procedings of UK planning and scheduling special interest group*, 1994.

D. Traum, S. Marsella, J. Gratch, J. Lee, and A. Hartholt. Multi-party, multi-issue, multi-strategy negotiation for multi-modal virtual agents. In *8th International Conference on Intelligent Virtual Agents.*,

D.M. Wegner. *The illusion of conscious will*. MIT Press, 2002.

H.M. Wellman, D. Cross, and J. Watson. Meta-analysis of theory-of-mind development: The truth about false belief. *Child development*, 72(3):655–684, 2001.

T. Wharton. *Pragmatics and the showing–saying distinction*. PhD thesis, UCL, 2003.

L. Wittgenstein. *Philosophical investigations, trans. G.E.M. Anscombe*. Oxford: Blackwell, 1953.