Reporting, Dialogue, and the Role of Grammar

Eleni Gregoromichelaki and Ruth Kempson

1 Introduction

There is a lot of debate in the literature as to whether metalinguistic, echoing or metarepresentational phenomena require semantic or pragmatic explanations or, perhaps the widest consensus, a mixture of the two. Recently some attention has been paid on whether grammatical models, i.e., models that define syntactic-semantic mappings (see e.g. Potts 2007; Ginzburg and Cooper 2014; Maier 2014), can offer a more substantial contribution in answering this question. In this chapter, we argue that they can, but not under standard assumptions as to what kind of mechanism "syntax" is and what the differentiation is between grammatical and pragmatic processes. Like Ginzburg and Cooper (2014) we take natural languages (NLs) to be primarily means of social engagement and on this basis we believe that various mechanisms that have been employed in the analysis of conversation can be extended to account for metarepresentational phenomena, which, as stressed in the Bakhtinian literature, demonstrate how dialogic interaction can be embedded within a single clause. However, we take such phenomena as a case study to show that a model adequate for accounting for the whole range of metalinguistic data, as well as for their interaction with other dialogue phenomena, has to depart from some standard assumptions in grammatical theorising: (a) we have to abandon the view of syntax as a separate representational level for strings of words, and (b) we need to incorporate in the grammar formalism various aspects of psycholinguistic accounts of NL-processing, like the intrinsic incrementality-predictivity

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

A. Capone et al. (eds.), *Indirect Reports and Pragmatics*, Perspectives in Pragmatics, Philosophy & Psychology 5, DOI 10.1007/978-3-319-21395-8_7

of parsing/production, and a realistic modelling of the context as information states that record or invoke utterance events and their modal and spatiotemporal coordinates.

The structure of this chapter is as follows: Firstly we present the traditional distinction between direct and indirect discourse (Sect. 2). We then show that the echoing and metarepresentational abilities that underpin such uses are not peculiar to reported discourse but occur also in dialogue, in particular in cases of repair and the process of "grounding", i.e. signalling of comprehension, (dis)agreement, or request for clarification of a previous utterance (Sect. 3). We then present a recent model, Ginzburg and Cooper (2014), which attempts to integrate the traditional view of direct/indirect discourse within a dialogue model without ad hoc devices not needed independently in the analysis of conversation (Sect. 4). We will then argue that the traditional direct/indirect discourse distinction cannot be maintained in view of various intermediate phenomena like free (in)direct discourse, hybrid, and mixed quotation (Sect. 5). In view of this we present an alternative grammar formalism that integrates some of the ideas of the Ginzburg & Cooper account but within a distinct incremental processing architecture that accounts naturally for the properties of these intermediate phenomena (Sect. 6).

2 Direct vs. Indirect Discourse

It has been assumed that talking about language or thought can be achieved in two distinct ways: directly or indirectly. In so-called "direct discourse"[1] one "pretends" to be the original speaker and reports the words accordingly; in "indirect discourse" it has been assumed that one reports the content of somebody's words or thoughts from one's own perspective. The first characterisation is akin to what in the Bakhtinian literature is characterised as the *pictorial style*, which emphasises the dialogicity of discourse by maintaining the style of the reported event; the second can be taken to correspond with the *linear style*, which, supposedly, focuses on the content of the reported speech and maintains a clear-cut boundary between reported speech and reporting context.

It has ten been argued that direct and indirect reports are grammatically distinct ways of reporting. In *direct discourse* the so-called framing verb combines with a clause which appears relatively independent, is attributed to the subject of the framing verb (the reportee), and there is a requirement for faithful or "verbatim" reporting (Maier 2015):

[1]We use the term *discourse* as neutral between reporting language (written or spoken) and thought.

(1) (Ann said:) "Really, I could care less about that" (said Ann).
[adapted from Maier 2014]

(2) She walked up to him and kissed him. "What am I doing? He is going to hate me now," she thought. [from Maier 2015]

In *indirect discourse* the framing verb combines with a subordinated clause, sometimes introduced with the complementiser *that*, whose content is presented from the reporter's perspective so that paraphrase is allowed:

(3) Ann said that she couldn't care less about that.

In the literature, direct reports are classified under *quotation*, i.e., the complement of the framing verb involves reference to linguistic objects. Indirect reports are analysed in parallel with intensional constructions so that the complement of the framing verb is taken to denote a proposition (Kaplan 2007; Maier 2007). Various further syntactic/semantic characteristics are assumed to distinguish the two types (see e.g. Partee 1973; Maier 2015). In writing, direct reports are supposed to be enclosed in quotation marks while in spoken language intonation distinguishes the reported part (Potts 2007). Indirect reports may involve the presence of a special complementiser like *that* in English. In some languages, like English, there are also mood-indicator shifts in indirect reports. For example, questions are assumed to be reported with non-inverted word order and without the presence of auxiliaries. In indirect reports, because the object reported is taken to be a proposition, non-propositional speech-act indicators like exclamations, vocatives or imperatives do not constitute appropriate markings on the verb's complement. It is also assumed that due to the independence of the reported clause in direct reports various syntactic processes are blocked for example, *wh*-extraction is not allowed. Schlenker (2011) argues that grammatical dependencies cannot "cross quotation marks" which accounts for the fact that (4) below has only the reading in (a), not the one in (b) which would be the one obtained if the embedded sentence were construed as a direct report:

(4) What did John say I ate?
(a) Indexical pronoun interpreted in current context: 'What did John say I (= the current speaker) ate?'
(b) *Shifted Reading: 'What did John say he (= John) ate?'

This judgement relies on the fact that, semantically, the interpretation of indexical expressions, like *I*, *you*, *here* and *tomorrow* in indirect reports is assumed to depend on the reporting utterance's context while in direct discourse it follows the reportee's context. Additionally, any tense occurring in the embedded clause of an indirect report follows language-particular sequence-of-tense rules.

However, there are semantic commonalities in the two types of report. Firstly, as has been pointed out since Partee (1973), the supposed independent sentence in direct reports is actually interpreted (contrary to the assumption that it constitutes pure "metalinguistic" use), witness a host of anaphoric phenomena, e.g. pronoun reference and ellipsis, that can depend on this independent sentence's content:

(5) "I talk better English than the both of youse!" shouted Charles, thereby convincing me that he didn't. [from Partee 1973]

(6) The sign says, "George Washington slept here", but I don't think he ever did. [from Partee 1973]

Further, as is well-known, indirect reports, despite the supposed current-speaker's context perspective, block logical entailments that are encountered in "transparent" environments. This has led to analyses of indirect reports as involving two separate sentences, for example, (7) has the underlying structure of (8) (Davidson 1968):

(7) Galileo said that the earth moves

(8) The earth moves. Galileo said that.

However, it is clear that this contradicts the syntactic criteria offered for the differentiation of direct from indirect reports, for example, the fact that blocking of *wh*-extraction is due to the presence of two independent clauses in direct reports. Such an analysis can then be put forward as an argument for an independent level of syntactic analysis, independent from semantics/pragmatics, with its own constraints and restrictions. Such an assumption is standard now in all formal approaches to NL-analysis. However, in our view, it can only be maintained by arbitrarily restricting the domain of data that constitute the remit of NL-grammars, in particular, by excluding data of NL-use, for example, the elliptical, fragmentary and incremental nature of structures and interpretations occurring in everyday conversation, including phenomena dismissed as "disfluencies", and non-linguistic behaviours. When we take such a broader view of NL, imposing the requirement on the grammatical formalism to be able to account for all such human-interaction data, we argue that direct and indirect discourse phenomena can be seen from a new perspective that unifies them under a holistic model of NL-use. We will now move to examine what unites all these traditionally considered distinct phenomena.

3 Mentioning Devices in Dialogue

Standard philosophical and linguistic approaches to the analysis of direct/indirect discourse take the view that these are idiosyncratic forms of language use, with particular formal characteristics and semantic entities involved. In fact, even approaches that take a dialogic perspective in the analysis of NL dismiss the prospect of a unified account of both conversational phenomena and reporting structures:

> "We are dealing here with words reacting on words. However, this phenomenon is distinctly and fundamentally *different from dialogue*. In dialogue, the lines of the individual participants are *grammatically disconnected*; they are not integrated into one unified context. Indeed, how could they be? *There are no syntactic forms with which to build a unity of dialogue*." (Volosinov, 1929/1973: 63; emphasis ours)

This perspective has been recently challenged in the domain of formal and psycholinguistic NL-models. In the domain of semantics/pragmatics such formal models do not necessarily restrict their remit to the level of single sentences. Instead, several researchers are currently interested in modelling the capacities underpinning NL use (see Cummings, this volume, Wieland, this volume) rather than formulating abstract systems that enforce standard competence-performance distinctions. In the domain of formal semantics, this has led to border disputes with pragmatics in that it's no longer clear whether the separation between "linguistically encoded" meanings and online, dynamically-derived ones, in interaction with contextual factors, can be enforced. In this respect, there has long been work emphasising the role of linguistic underspecification in the process of deriving meaning in context (see, e.g., Sperber and Wilson 1995; Levinson 2000; Recanati 2004, 2010) and formulating notions of 'procedural meaning' that cannot be accommodated under truth-theoretic conceptions of semantics (e.g. Blakemore 1987). Further inadequacies of traditional truth-based theories and the focus on single sentences have been highlighted by the "dynamic turn" in semantics (DRT, Kamp 1981; Kamp and Reyle 1993; DPL, Groenendijk and Stokhof 1991 and related frameworks) that have drawn attention to the importance of conceiving meaning as updates to 'information states' rather than, statically, as sets of truth conditions (propositions) assigned to sentences. Nevertheless, all these approaches still concentrate on individual mental states modelled as autonomous representations that abstract away from the social and material circumstances of NL processing.

3.1 Dialogue Phenomena: Echoing and Grounding

On a more psychologically-realistic perspective that looks at NL in terms of its use as part of social cognition, reported speech appears as only one aspect of a general phenomenon that is regularly encountered in the analysis of everyday conversation. First of all, it has been argued that fundamental processes in human interaction, the mechanisms explicating the contribution of non-declarative moods, employ metarepresentation of the embedded proposition (Wilson and Sperber 1988) or encode instructions on how to accommodate these contents in the discourse model (Gregoromichelaki and Kempson 2015; Eshghi et al. 2015). Further, in contrast to views that take NLs as abstract formal systems dealing with the definition and interpretation of well-formed sentences, some recent systems modelling NL use propose grammars that have the capability to offer syntactic mechanisms that can cross, not only sentential boundaries, but also turn boundaries (see e.g. Gregoromichelaki et al. 2011), as is required in order to provide natural and motivated accounts of phenomena where the construction, interpretation and authorship of utterances is spread across participants (*split utterances* see turns 3, 4, 5, 12, 14, 21 below):

(9)

1. A: Instead of having < *name hidden* > <*unclear* > they had to come through the Dock Commission all of the men, they wanted so and so men for that boat, they used to come through to me.
2. B: Before that though, <*name hidden* > and < *name hidden* > [<*unclear* > had their own men]
3. A: [Had their own men
4. B: unload the boats?
5. A: unload the boats, yes. They < *unclear*>
6. B: They were employed directly by
7. A: That's right but they all came
8. B: < *name hidden* > ?
9. A: They used to work say 1 week and have about a month off or go on the dole for a month.
10. B: So then what happened was, did the Dock Commission say you can't have your own men anymore?
11. A: That's right they had to go on a rota.
12. B: Run by the Dock Commission?
13. A: Run by the Dock Commission. See the dockers then all got together and they said right so many men for that job, so many for that job and that didn't matter who they were, they had to < *unclear* > their job, all the way round the dock.
14. B: Whether they wanted to go on that job or not?
15. A: Whether they want to go or not, they take their turn and the employer had to pay a percentage into the pool what those men earned, so when those men hadn't work at all they drew their money from the National Dock Labour Board.
16. B: Is this where the National Dock Labour Board came into existence?
17. A: That's how how they come into existence, yes < *name hidden* > he was a man what introduced that.
18. B: When was this?
19. A: Oh that's er, I would say about nineteen forty roughly < *clears throat* > Id say about nineteen forty that came in, might have been before that.
20. B: Before that then if they were ill
21. A: They get nothing.
22. B: Could they not get any welfare benefit?
23. A: No [BNC, H5H: 89–113]

In many cases of split utterances, as in reported discourse, the apparent speaker can be seen as the *animator* but not necessarily the *author* or *principal* (Goffman 1979; Antaki et al. 1996). For example, in (9)-4 and (9)-12 above, the continuations are offered by interlocutor B accompanied by a request for confirmation towards A as to whether they reflect A's view of the situation, i.e., whether they provide contents that the actual principal, A, deems as appropriate or whether they are an appropriate "echoing" of A's authorship, i.e., what they were going to say. Similarly, during interaction, issues of perception or interpretation of the interlocutors' words constantly arise and various strategies for the resolution of such issues are employed (as can be seen in (9)-12, (9)-14, (9)-18 etc.). These mechanisms rely in the current speaker's utterance depending on the articulation of the previous speaker's

utterance, and, sometimes, employing its form and content as antecedents, explicitly or covertly, a characteristic in common with reported discourse, but transcending turn and interlocutor boundaries. Such phenomena have been analysed as requiring both incremental licensing of NL structures and contents but also as relying crucially in the intrinsically predictive nature of NL processing (Gregoromichelaki 2013a, b). In fact, it has been argued that reported speech is one of the environments where the phenomenon of split-utterances is observed frequently due to the assumed projectibility of the upcoming continuation (Lerner 1991) and, we would add, the potential for open authorship and assumption of responsibility for the speech act performed by such structures:

(10) A: mid April. we had reached the point of thinking that we weren't going to be able to reach a policy decision
B: that's right
A: and so we must. tell these guys [that we'll carry on ..]
B: [we're going to carry on. yep] [from Antaki et al. 1996]

(11) Anne: I wish that he'd say- he said, "I have to be back around four because our family is having something," and I wish he'd say
Kay: "why don't you come over honey"
Anne: Yeah. [from Lerner 1991]

(12) Ken: she'll say // wouldja-
Louise: wanna glassa milk? // hehhh
Ken: No. wouldju like a little bitta he'ing?
Louise: heh// ha ha
Ken: wouldja like some crekles?
Louise: ehh ha ha ha ha
Ken: wouldja like a peanut butter an' jelly sandwich? [from Lerner 1991]

(13) Roger: they rationalized it. they say heh heh heh
Louise: it wasn't there it was a(h)II in hi(h)s imagination. [from Lerner 1991]

From this perspective, "quotation" mechanisms are crucially involved in the functioning of dialogical interaction in that every utterance responds to an antecedent one and is construed as backward-looking commentary on that antecedent utterance (see e.g. *grounding*, Clark 1996, Ch. 8; Schegloff 2007) as well as a forward-looking action anticipating a response (Gregoromichelaki 2013a, b; Arundale and Good 2002). These mechanisms appear more transparently on the surface in cases of metacommunicative exchanges. This is more clearly shown when "echoing" means are used, for example, in cases of *clarification questions* which include reference to some utterance token (Ginzburg 2012):

(14) A: Did Bo leave?
B: Bo? ('Who are you referring to as 'Bo'?'; 'Did you utter the word *Bo*?')

(15) A: Who came?
B: Who came? How dare you? ('Are you asking "who came"?')

Notice that such phenomena, even though they display "metalinguistic" features, normally pattern syntactically with indirect reports in that indexicals take their reference from the parameters of the current context as it shifts incrementally during the unfolding of the utterance (Gregoromichelaki et al. 2011):

(16) A: Oh, I am so sorry, did you burn
B: myself? No, it's ok

(17) A: Did you leave?
B: Me? ('Are you asking about ['the current speaker'] ?')

However, the responsibility for the utterance act in question is not necessarily the current speaker's, as we said the *principal*, i.e. the agent of the utterance event is not necessarily identified with the agent of the speech act (Gregoromichelaki and Kempson 2015). So the agent continuing another's utterance can question the previous speaker as to why they are performing the relevant speech act, a practice normally infelicitous (Ginzburg 2012) in circumstances where the two roles (agent of utterance, agent of speech act) are identified:

(18) A to C: Didn't B say yesterday that he's paranoid. Why?
'Why am I asking if B said B is paranoid'

(19) A to C: Didn't B say yesterday that . . .
B to C/A: I am paranoid? 'B is paranoid'
B to A: Why? 'Why is A asking C whether B said that B is paranoid?'
A to B: Because you told us five minutes ago you're easy-going.

In (19) the agent of the questioning speech act 'Didn't B say yesterday that B is paranoid' is taken to be A, hence the felicitous *why*-questioning of A by B. Yet, the string *Didn't B say yesterday that I am paranoid* gives a misleading interpretation, if it's considered as a continuous string of words encoding a speech act performed by A, namely 'Didn't B say that A is paranoid'. Indexicals in indirect reports are supposed to conform to the current contextual parameters, which is indeed what happens in (19). However, because the context shifts during the reporting, the agent of the embedded assertion, B, and the agent of the relevant part of the utterance coincide temporarily, which, when considered from a global sentential point of view gives a misleading form. This is confirmed by a direct report version:

(20) A to C: Didn't B say yesterday . . .
B to C: "I am paranoid"? '
B to A: Why? 'Why is A asking C whether B said that B is paranoid?'
A to B: Because you told us five minutes ago you're easy-going.

In this case the global string *Didn't B say yesterday "I am paranoid"?* gets accidentally the correct interpretation. Expressed in more standard formal semantic terms, this shows that interpretations for such strings cannot be derived globally but instead have to be derived incrementally at each word stage, with each word's *character* (Kaplan 1989) interacting with the contextual parameters at each subsentential stage of utterance. The final stage of semantic composition is then

only dealing with contents rather than sentential characters. Moreover, in terms of pragmatics, it shows that issues of speech act responsibility (agency) have to be determined separately from fixing the contextual parameter of 'speaker' (utterer), which, as we will see below, is a mechanism needed for accounting for cases of mixed quotation.

Asking an elliptical *why*-question as in (18)–(20) above, makes implicit reference to a particular antecedent utterance event and asks for its purpose. In (14)–(15) particular phonological tokens are repeated and have to be recognised as such, i.e. as repetitions, echoes, of the antecedent tokens in order to be interpreted. In addition, phenomena of *other-* or *self-repair* also require mechanisms that enable the recognition of a new token being produced as similar in form to an antecedent one, and then another token being offered as a replacement in terms of content for that antecedent one:

(21) A: Bo, (not Bo,)(I mean) Joe, left
(22) A: Bo left. (Not Bo,)(I mean) Joe.
(23) A: Bo
B: (Not Bo)(You mean) Joe.
A: Yes. He left
(24) A: Bo left.
B: (Not Bo) (You mean) Joe.
A: Yes.

Given that all these phenomena, split-utterances, clarification, and repair exchanges, are initiated from very early on in language acquisition, the means and skills involved in the production/comprehension of reported discourse do not appear so idiosyncratic (cf. Wieland, this volume). Taking this assumption seriously, two recent holistic models of NL use, HPSG-TTR (Ginzburg and Cooper 2014) and Dynamic Syntax-TTR (Gregoromichelaki, to appear) seek to model reported discourse via the same mechanisms as those used to analyse everyday conversation. We turn to these two models next.

4 Ginzburg and Cooper (2014)

4.1 Utterance Events and TTR Representations

4.1.1 Utterance Events in DRT

Recent efforts in formal semantics, inspired by work in Situation Semantics and DRT, have shifted attention away from a strict formulation of a truth theory for sentences as a theory of semantic competence to developing theories of semantic interpretation for utterances in context. For this purpose, a common representational system allowing the specification and seamless integration of multiple types of

information has been sought. One strand of this development, based on recent advances in developing compositional forms of DRT, is the PTT model (Poesio and Traum 1997, 1998; Poesio and Rieser 2010), which expands the dynamic view of semantics to take into account underspecification of meaning resolved in context and NL use in interaction. One distinctive feature of Poesio and Rieser (2010) is the assumption—derived from ideas developed in Situation Semantics (Barwise and Perry 1983) and Clark (1996)—that the 'information state' (context) representation of each participant in a conversation, also includes the reification and explicit representation of the *utterance event/situation*, i.e., the contextual parameters of the conversation itself. So, along with the mutually accepted truth-evaluable content of utterances (*common ground*), information about the discourse situation is recorded in a unified representation, a discourse representation structure (DRS), modelling each participant's 'information state' at each point in the dialogue. Even more innovatively compared to previous versions of DRT, here the occurrence of utterances of sub-sentential constituents is recorded in this representation as the occurrence of events in a certain temporal order (*micro-conversational events*) which thus become part of the information state. The occurrence of each such micro-conversational event leads to immediate updates of the participants' information states with the initiation of semantic and pragmatic interpretation processes (Larsson and Traum 2000; Stone 2004) following the specifications of the grammar. As regards pragmatic integration, in this model, speech acts are conceptualised as events too, termed as *conversational events*, since just like any other events, they can serve as the antecedents of anaphoric expressions:

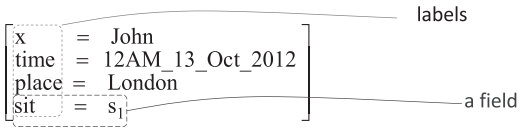
(25) A: You're an idiot.
B: *That* was uncalled for. [*that*: A insulting B]

4.1.2 Utterance Events in TTR

Another recent articulation of this effort has been via the development of Type Theory with Records (TTR). TTR provides a transparent semantic representation format that can integrate both low-level (sub-symbolic) perceptual information (see e.g. Larsson 2011) and underspecified, flexible meanings of NL expressions (see, e.g. Cooper 2005, 2012). Such integration allows the modelling of how NL forms and meanings adapt to the discourse situation via the formalisation of an evolving, structured notion of the (multi-modal) context. Consequently, instead of adopting the assumption that the role of semantic theories is to assign truth conditions to decontextualised *sentences*, in these approaches, attention has shifted to the modelling of situated *utterances* and speech acts. This has led to a significant expansion of the data deemed appropriate for inclusion in a formal theory of interpretation, namely, the modelling of the use of language in interaction and the demands that this places on appropriate semantic models (see e.g. Ginzburg and Cooper 2004; Ginzburg 2012).

TTR is a representation language that provides recursive data structures reminiscent both of HPSG type-feature structures and, semantically, of discourse representation structures (DRSs). Records, like the record *r* below, are structured collections of statements ("fields") consisting of assignments of entities to "labels", the equivalent of discourse referents/variables in DRT:

(26) r =



Such records can then be taken as the representation of events/situations in the world. Importantly, contexts of utterance and the actual speech events that take place within them are represented by such records too. Records (and therefore events/situations) are classified by types which are called *record types*. Unlike the basic Montagovian types, record types are structured and recursive (i.e. types can be embedded as the value of labels within types) and dependencies are allowed among the values assigned to the labels. A record *r* belongs to a type *T* iff each field in *r* satisfies the constraints specified by *T*. For example, as a simplified illustration, the record *r* in (26) is of the type *T* in (27) below (it is a *witness* for *T*) because *r* assigns entities to the labels that satisfy the type requirements specified by *T*. This means that the label *x* is assigned an entity of type IND(ividual), namely, John, the labels *place* and *time* are assigned entities that are places and times respectively and the event s_1 is such that it is of a type that indeed contains evidence that John runs – perhaps it is an observation or some actual event of John's running (this latter characterisation is related to Martin-Löf's "propositions as types" idea, hence RUN(JOHN) is a type here):

(27) $T = \begin{bmatrix} x & : & \text{IND} \\ time & : & \text{TIME} \\ place & : & \text{PLACE} \\ sit & : & \text{RUN (JOHN)} \end{bmatrix}$

Types, which can thus be conceived as categorisations of events and entities, are what provides the interface between the external world and cognition; for example, record types, namely, categorisations of situations, can be used to provide representations of perceptual judgements, meaning relations, grammatical information, speech act assignments, etc. In addition, in TTR, types are first-class citizens of the semantic ontology, not reducible to sets of their members. So types are intensional and inference can be performed at the level of types, irreducibly *about* the types themselves, solving puzzles that traditionally have been encountered in intensional constructions such as the complements of propositional attitude and reporting verbs. Moreover, because types are always modifiable by adding/deleting fields, the underspecification and subsequent enrichment that permeates type judgements during, e.g., language acquisition, knowledge adjustment, conversational coordination and even, as we will see now, quoting (some aspects of) another's speech, are naturally handled.

4.2 TTR-Modelling of Reporting Constructions

Currently a number of accounts have been proposed regarding the semantics of reported speech. However, such accounts devise ad hoc entities in order to shift the usual contents and otherwise ignore the contribution of the whole grammatical apparatus, e.g. syntax, phonology etc., and even pragmatics. In contrast, the account presented in Ginzburg and Cooper (2014) (G&C henceforth) provides syntactic analyses, denotations and pragmatic constraints for reporting constructions that utilise independently needed grammatical entities. G&C aim to demonstrate that a dialogical perspective on NL structure and use provides directly the tools to deal with reported discourse via structures and denotations that are already independently motivated for the modelling of dialogue phenomena.

4.2.1 Grounding and Clarification

Following the model most comprehensively detailed in Ginzburg (2012), the analysis of dialogue involves richly structured representations of context ('information states'). Adoption of the TTR formalism allows Ginzburg to construct models of the semantic ontology and the grammar as well as a model of how contexts evolve during the conversation. To account for the metacommunicative function of certain utterances, in particular, clarification requests (see (14) earlier), dialogue processing is assumed to crucially require *grounding* (Clark 1996), a process during which each participant either confirms that they have understood the utterance addressed to them, thus incorporating it in their information state, or seek clarification of aspects that have not been "grounded". Ginzburg extends the grounding requirement along two dimensions. Firstly, grounding is not immediate; instead, it allows partially comprehended utterances to contribute to the context while ungrounded (parts of) utterances can remain as "pending" and lead to metacommunicative interaction (clarification) for their resolution. Secondly, it is not only semantic content that is recorded in the participants' context, but also a range of properties of the utterance that has occurred, e.g., syntactic/phonological information that would enable the disambiguation and resolution of elliptical utterances that function metacommunicatively (see (14)-(17) earlier). These extensions require that the grammar should be able to express reference to utterances as "utterance events" specified along multiple dimensions. Records, like the one we saw in (26) earlier, are employed to serve this role. (Partial) Grounding is then formalised through the pairing of an utterance event (a record, a token) with a (partial) utterance type, i.e., a grammatical type (a "sign", recording grammatical aspects defined in HPSG) that classifies it. Such signs are modelled as record types, like the one shown in (27) earlier. So here a major advantage of the use of TTR becomes evident: the grammar and the conversational mechanisms are provided with access to both types and tokens of utterances, which forms the basis for modelling metacommunicative or metalinguistic functions of NL elements. For example, it is argued that the clarification request in (28) below, which echoes A's use of *Bo*, has a reading which

queries which individual named "Bo" the speaker was referring to in the previous utterance ("intended content reading"), not who the name *Bo* refers to in general – which is also a possible reading, as can be seen more clearly in the "intended content reading" of the predicate 'finagle' in (28). These readings need to be disambiguated from other readings such as the "clausal confirmation" readings in (c):

(28) a. A: Did Bo finagle a raise?
B: (i) Bo? / (ii) finagle?
b. Intended content readings:
(i) 'Who is (the) "Bo" (you're referring to)?' (ii) 'What does it mean "to finagle"?'
c. Clausal confirmation readings:
(i)'Are you asking if BO (of all people) finagled a raise?' /
(ii) 'Bo FINAGLED a raise (of all actions)?'

If the grammar and the model of the participants' information states allow for reference to actual token utterance events, it becomes possible to explicitly model readings such as that in (28)b. By assigning interpretations to the fragment *Bo* that match the intuitive paraphrase given involving reference to the specific utterance event that has occurred, namely, A's uttering *Bo*. In order to achieve this, the model is formulated in a constructional version of HPSG expressed in the representational framework of TTR. The rich type theory included in this model allows for the definition of entities that the grammar and the model of the context can manipulate both at the level of utterance tokens (events) and utterance types (signs).

4.2.2 Locutionary Propositions and Abstract Semantic Objects

Two components in this process of grounding are relevant for the analysis of direct/indirect discourse that concerns us here: (a) locutionary propositions, and (b) abstract quasi-propositional objects assigned as contents to sentential units in order to serve as the arguments of speech act predicates. For an utterance in dialogue to be grounded, first it has to be parsed and understood correctly. The outcome of this process of parsing is modelled via requiring the truth of a so-called *locutionary proposition*. Simplifying somewhat, a locutionary proposition is the pairing of the current utterance event token with a fully-specified grammatical type (an HPSG-defined "sign"). Such signs are structured types, i.e., representations that include domains for phonology, syntax, semantics and pragmatic specifications with constraints governing their correspondence. If the truth of such a locutionary proposition cannot be established after parsing, i.e., if a complete grammatical type cannot be assigned to an utterance, various clarifications are licensed to occur that can make reference to the particular utterance token that causes the trouble. During this process, reference can also be made to the particular speech act performed by the previous interlocutor, e.g., modelling interpretations like 'Are you asking q', 'Are you asserting p?'. The quasi-propositional arguments *p* or *q* in such speech act specifications indicate abstract semantic objects like propositions, questions, outcomes, facts etc. These objects are defined in the semantic ontology and are assigned by the grammar as the contents derived through the realisation of the

speech acts performed with utterances. For example, root clauses are required by the grammar to include a speech act specification, selected from a small number of such specifications, like *Assert*, *Ask*, *Order*, *Exclaim*. Which of these speech act specifications is selected depends on the semantic object that such specifications attribute to the agent of the speech act. So, a *proposition* will be what an agent *Assert*s, a *question* will be the complement of the *Ask* relation, an *outcome* the complement of the *Order* relation, and a *fact* is the object associated with the *Exclaim* relation.

These abstract entities, locutionary proposition and abstract quasi-propositional semantic objects, that have been postulated independently for the licensing of NL use in conversation, especially metacommunicative interaction as in (14), are taken by G&C to naturally extend to pure and direct quotation, and indirect reports.

4.2.3 Pure Quotation

Having assumed a constructional version of HPSG, in applying these mechanisms to reporting discourse, G&C define *constructions* for various quotation phenomena that specify the syntactic, semantic and pragmatic constraints for their licensing. According to various theories of quotation (*identity theory*, Washington (1992); Saka (1998, 2011); *proper name theory*, Tarski (1933); Quine (1940), and *description theory*, Geach (1957), see Cappelen and Lepore 1997), the quotation marks modify the reference of the expression that is enclosed within them, so that the expression now refers to itself, rather than its usual denotation, an entity in the world. However, as C&G argue, pure quotation cannot be taken as simply involving reference to "expressions" in general. Their model offers a specification of the notion "expression" via the resources of some grammar that includes a characterisation of the TTR type of such token expressions (see also Maier 2014; Potts 2007). So, in analysing pure quotation, G&C assume that the contextual parameters included in the information state have to include a parameter Γ that refers to the grammar that licenses the type of the expression used. Γ roughly corresponds to what in Recanati (2010) is characterised as a "language", or as we will prefer to define it, for reasons that we will explain below, a conceptualisation of NL-use that reflects folk-linguistic conceptions but does not necessarily correspond with the analysts' grammar of a particular language (unless of course the discourse involves discussion of exactly such a grammar). Under this assumption, pure quotation, which is usually assumed to introduce a referential term that refers to the linguistic material enclosed in the quotation marks, is accounted for in G&C via a construction which licenses an NL term, e.g. *Mary*, to syntactically project a phrase whose semantic content is some aspect of the grammatical type (sign) relative to a particular grammar Γ, e.g., its phonology in a case like:

(29) 'Mary' starts with 'm'

The contextual parameters usually assigned by Γ to standard uses of the sign (e.g. speaker-hearer, time, location etc.) are discarded in such a quotational construction. This explains the opacity of such uses. Use of quotation marks in written discourse indicates this shift of content for such uses of signs.

4.2.4 Direct and Indirect Quotation

Turning to the analysis of direct and indirect discourse, these are seen as involving two components: (a) lexical entries for the framing verbs ('quotative predicates') and (b) constructions that specify the presumed idiosyncratic properties of such structures. Quotative predicates select for clauses denoting either (a) locutionary propositions, or (b) quasi-propositional abstract entities (see earlier Sect. 4.2.2). Both indirect reports and direct quotation are analysed as constructions that involve the combination of a framing verb, like *say* or *ask*, with a sentence whose denotation involves a quasi-propositional abstract semantic object.

A *direct-quotation construction* involves, firstly, the projection of a *direct-quotative phrase* from the quoted material. This phrase will then serve as the complement of a framing verb specified to require (the supertype of) such a complement. The derived semantic content of a direct-quotative phrase is a locutionary proposition, i.e., an utterance event to which a grammatical type, a "sign", is assigned by a particular grammar Γ (see earlier Sect. 4.2.2). However, unlike standard utterances in dialogue, the event component of such a proposition becomes the reported utterance event (simplifying somewhat) rather than the current reporting utterance event. This reported event though is now associated with a grammatical type assigned to it by the reporter relative to a grammar Γ thus accounting for the fact that, for example, the quote might be in a language different than the one of the original reported event or other modifications the reporter might effect (and still be counted as direct quotation). Since the grammar manipulates TTR types as well as tokens, it now becomes possible to express how the original utterance event and the reporting event are deemed to be "similar" in some respects (see e.g. Clark and Gerrig 1990). The grammatical type assigned to the reporting event by the reporter's assumed grammar Γ is constrained to "resemble" the type of the original event, i.e., there has to exist a contextually-defined value on a similarity measure between the grammatical types of the original and the reporting events. Further, even though via this construction the contextual parameters of the standard use of the sign are discarded, as we also saw in pure quotation cases earlier, for direct quotation, at the phrasal level, a new set of contextual parameters is introduced via the representation of the original utterance event and its grammatical type. In this way, the content of the reported sign becomes available. This allows for the explanation of cases of anaphoric reference to the *content* of the quotation subsequently, as in the Partee examples in (6) earlier. A further innovative advantage offered by this analysis is that by analyzing direct quotation complements as denoting locutionary propositions, which include as one of their component a "sign" (a grammatical type) we can explain the fact a single sentence can contain predications that make use of both

type and token aspects of a quotation, e.g., use the same quotation as both the complement of direct-quotation construction and as a pure quotation, as in (30)-(31):

(30) 'Was I snoring' *was asked by Bill* and *is a frequently used interrogative clause*.
(31) 'Am I snoring?' *asked Bill*, *a sentence* frequently uttered by men who don't think they snore. It is usually answered by 'You were before you woke up.'

The direct-quotative phrases whose properties we have just described are the complements of verbs that combine with direct quotations, i.e., independent clauses. Many such verbs also take embedded clauses as their complements, resulting in indirect discourse constructions. This is implemented in this model by defining such verbs to combine with complements that can have two distinct semantic objects as contents. For example, the lexical entry for *ask* has two versions. In the case of direct quotation, the lexical entry for the verb *ask* specifies that the complement must have as its content a locutionary proposition, i.e., the combination of an utterance event with a grammatical type. As we've just seen, due to the direct-quotative phrase specifications, the utterance event will be the reported event (simplifying somewhat) and the grammatical type a type similar to the one assigned to this reported event. Additionally, the lexical entry for the verb *ask* specifies that the SPEAKER *x* of the utterance event included in this locutionary proposition (the original utterer) is identified with the subject of the main clause. Since the grammar, according to Ginzburg (2012), conventionally associates speech act specifications with utterances (see earlier Sect. 4.2.2), the speech act characterisation of the original reported event is available through the usual grammatical type associated with it. Accordingly, the content of the main clause is inherited from the grammatical type of the complement of *ask* so that it comes out as the speech act specification *Ask(x, q)* where *q* is an abstract semantic object of type *question*. (Note that to this a new speech act specification will be added to the effect that the final content will come out as *Assert(Speaker, (Ask(x,q))*).

On the other hand, the lexical entry for *ask* in an indirect-report context specifies that it combines with a subject *x* and a sentential complement. Unlike the case with direct quotation, this sentential complement is NOT of the type 'locutionary proposition', i.e., the original reported event is not included in the representation, hence, unlike direct quotation, it cannot affect the contextual parameters. The only restriction here is that the complement has as its content an abstract semantic object *q* of type *question* (e.g. *whether John left*). The content of the whole sentence built on the basis of the lexical entry for *ask* is then a proposition *Ask(x,q)* where *x* is identified as both the subject *x* of the main clause and as the agent of the speech act reported (the eventual content derived will again be *Assert(Speaker, (Ask(x,q))*).

This account is designed to capture the commonalities of direct/indirect discourse via the lexical entries of verbs that combine with both. As we just saw the contents derived for both such structures are identical, even though the structure with the direct-quotative phrase includes reference to the original demonstrated event. Another commonality this setup is designed to capture is the common entailments between direct and indirect reports, illustrated by the fact that they both support common inferences about the characterisation of the semantic object they combine

with. So both (32) and (33) below entail (34), which is explained because, as we just saw, the contents assigned to the sentences built on the basis of the two versions of *ask* are identical:

(32) Zohar asked whether she snored.
(33) Zohar asked 'naxarti?'.
(34) Zohar asked a *question*, a question about herself.

So the G&C captures successfully various properties of reporting constructions via the attempted unification of the mechanisms of quotation with mechanisms of repair in conversation. In addition, the TTR modelling proposed is able to allow for the explanation of new data like the cases of "mixed predications" in (30)-(31) where a single predication can address simultaneously both metalinguistic and reporting aspects of the same utterance. It also claims to capture the commonalities between indirect and direct reporting and the common and mutual entailments holding between such structures as seen in (32)-(34) earlier.

We believe that there are some problems with this latter claim, stemming from the fact that the grammar associates conventional speech-act specifications with each main clause. For the same reason, in combination with the fact that the grammar is defined in terms of constructions, rather than general structural constraints, the account does not seem to be able to generalise to cover all quotational possibilities that have been reported in the literature. We will argue below that the main technical and conceptual reasons for this are, firstly, the fact that the intrinsic incrementality of NL-processing is not part of the grammar, and, secondly, the fact that syntax is taken as an independent level of analysis with its own categories and constraints (as is standard for most grammar formalisms). In order to remedy these shortcoming we will then propose an alternative account that builds on G&C but within an incremental, dynamic framework.

5 Free (In)direct Discourse, Mixed Quotation, Hybrid Uses

Recanati (2001) makes a distinction between closed and open quotation. *Closed quotation* are instances where the quotation semantically plays the role of a singular term and fills a syntactic slot in the sentence. The C&G account is explicitly addressed to such closed quotation cases only. However, we believe that G&C have provided some of the resources that make a more inclusive account available, i.e., dealing with the phenomenon of *open quotation*, where the quoted material fulfils its usual role in the sentence, or none at all if it is a main clause. The only factor that prevents an easy integration of such phenomena in the G&C account is, in our view, the standard assumption of an independent syntactic level of analysis in the grammar and the lack of an incremental syntactic licensing and interpretation. The same assumptions, standard in all formal grammatical frameworks, prevent other grammatical accounts of quotation (e.g. Potts 2007; Maier 2014) to deal with the whole range of data as we will show now.

The C&G constructional account inevitably adheres to the standard strict division between direct and indirect quotation. However, these strict distinctions have been disputed (see e.g. Allan, this volume; Holt, this volume) as there is a host of phenomena that lie in a continuum between these two supposed extremes. Firstly, there is always the possibility to introduce quotative elements, for example, elements extraneous to the reporter's dialect, within a report otherwise characterisable as indirect (and without the use of quotation marks, contra Maier 2014):

(35) To which Mr Bailey modestly replied that he hoped he knowed wot o'clock it wos in gineral. [Dickens, Martin Chuzzlewit, from Clark and Gerrig 1990: 791]

Then there is the phenomenon of *free direct discourse*. In these cases there is no framing verb or clause to indicate reporting but indexicals and other devices conform to the reported context indicating direct reporting:

(36) **Hilary** crept into the back room. **She** saw the curtains, dragged together roughly, as if – as if – There's someone behind them. **I**'m sure there's someone behind them. **I** must stay calm. – **She** reached for the light. [from Crystal 2013]

Free indirect discourse is similar to indirect reporting in that there is potential shift of tenses and indexicals. However, usually there are no overt reporting indications and some features of direct discourse (such as direct questions and vocatives) are maintained so that there is only a partial shift of perspective towards the reportee:

(37) **Mary** felt relieved. If Peter came **tomorrow**, **she** would be saved. [from Recanati 2000]

(38) **John** is totally paranoid. Everybody spies on **him** or wants to kill **him**, including **his** own mother. [from Recanati 2000]

(39) **Marie** was wondering. Did **her** brother arrive? [from Bonami and Goddard 2008]

And there are further "hybrid" cases, for example, in English, the interrogative word order can sometimes be maintained in indirectly reported questions:

(40) The baritone was asked **what did** he think of Mrs Kearney's conduct. [from James Joyce, *Dubliners*, cited by McCloskey 2006 in Köder and Maier 2015: fn. 1]

These phenomena cannot be handled by the C&G account because their model requires that the grammar deals with phrasal constructions that specify either direct or indirect features. In all these cases, however, there is no necessity for a framing verb to determine the appearance of a quotation-like interpretation. Another phenomenon that is excluded for the same reasons is that of *mixed quotation*, a combination of direct and indirect discourse, characterized, in written language, by the use of quotation marks in the sentential complement of an indirect-report construction (see e.g. Cappelen and Lepore 1997):

(41) Alice said that life "is difficult to understand".

In these cases, in common with indirect reporting, the complement of the verb is a *that*-clause which "samesays", i.e., has the same content as, what the reported speaker said (Davidson 1968). As in direct reporting, there seems to be indication

that the referent of the subject of the framing verb used similar tokens as those appearing in the report ("sametokening"). Since both these aspects of the reports affect the truth-conditions of the sentence, they need to be accounted for by an adequate model of NL use (Recanati 2000; Potts 2007; Geurts and Maier 2005). However, certain alleged peculiarities of mixed quotation create problems for both syntactic/semantic and pragmatic accounts. First of all, like direct quotation, and, as we saw earlier in (35), even with indirect quotation proper, there is the possibility to switch not only the interpretation of indexicals but even language in the midst of reporting such quotations:

(42) Wright won't disclose how much the Nike deal is worth, saying only that 'they treat me well'. (The Face, September 93: 55) [from de Brabanter 2010]

(43) A doctor tells him [Gustave Flaubert] he is like a 'vieille femme hysterique'; he agrees. (TLS online, 18 December 1998) [from de Brabanter 2010]

Another issue that arises for formalisms that do not embrace the incrementality of processing in the grammar, attempting to characterise and interpret well-formed sentences, is the fact that the quotation-like interpretation might span multiple sentences or even within-sentence non-constituents:

(44) She replied, 'I live alone. My son lives alone too. We both prefer it that way'. [from Huddleston and Pullum 2002: 1026, cited in de Brabanter 2010]

(45) Also, he categorically stated that "there is no legal way of temporal extension of the Greek debt without this being regarded as a credit event. Therefore there is no way that it will be allowed to happen such a credit event in Greece because it would create negative impact on the whole system." [from Gregoromichelaki to appear]

(46) Writing that book, Doyle felt himself 'a slave to reality. I was just dying to write a big book, and to have a bit of fun'. (Independent Arts, 17 September 2004) [from de Brabanter 2010]

(47) David said that he had donated "largish sums, to several benign institutions". [from Abbot 2005]

(48) Mary allowed as how her dog ate "odd things, when left to his own devices". [from Abbot 2005]

(49) Tim Marlow of London's White Cube gallery suggested that such self-censorship was now common, though 'very few people have explicitly admitted' it. (www.guardian.co.uk/commentisfree/2008/oct/01/religion.islam) [de Brabanter 2010]

(50) [The doctors'] actions defied the instructions of members of Congress, who issued subpoenas to attempt to block 'the barbaric' removal of her feeding tube on Friday [. . .]. (The Guardian online, 20 March 2005) [de Brabanter 2010]

This cannot be handled by a grammar that requires phrases to be built out of conventional constituents that just switch interpretation. As de Brabanter (2010) argues, the whole set of these effects cannot even be handled by the ad-hoc constituency imposed by Maier's (2007) account since the continuity and unity of the quoted fragments gets lost. And, as Recanati (2010), among others, point out

such phenomena have truth-conditional effects, as can be seen from the distinct interpretations obtained when the quotation marks are removed:

(51) Paul says he's due to present his work in the 'paper session' [Paul calls "paper session" the "poster session"]
(52) Paul says he's due to present his work in the paper session [Recanati 2010]
(53) James says that 'Quine' wants to speak to us [James thinks that McPherson is Quine]
(54) James says that Quine wants to speak to us [Recanati 2010]
(55) Nicola believes that his father is a 'philtosopher'
(56) Nicola believes that his father is a philtosopher [Cappelen and Lepore 1997]

In our view, this clearly indicates that a grammar formalism needs to integrate interaction with pragmatics at a subsentential level, before the semantic contents derived from words are composed. The cases above have been analysed by Recanati (2010) in terms of a language-shift. We can implement this, as in the G&H account, by assuming that one of the contextual parameters that need to be included in a grammatical analysis must be variables representing various entities like potential idiolects, dialects, languages etc. However these are factors that can shift during the interpretation of a fragment of the utterance being processed and do not necessarily either project syntactically or are defined at the root level. Moreover, we also believe that we need a rather liberal characterisation of the entity that represents such folk-linguistic assumptions since such metalinguistic characterisations are open-ended and are not dependent on any actual grammatical characterisation as the examples in (51)-(56) show.

Another issue that arises for the C&G account is the fact that the speech-act specification associated with each main clause is taken to be conventionalised, i.e. there is a selection from among a predefined set of such illocutionary forces (see earlier Sect. 4.2.2). However, we believe that the precise speech act specification potentially assigned to each utterance is open-ended and subject to pragmatic inference so that there can't be any default specifications determined by the grammar; the grammar just needs to offer the potential for such optional pragmatic inferences to affect truth-conditional content on the way to deriving contents for the full utterance. This is shown by the fact that indirect report complements can appear with a multitude of speech-act denoting framing verbs:

(57) Replying to another question by the shareholders he characterised as "imaginary scenario" the possibility of Greece leaving the eurozone, however, he clarified that "there is no practice or methodology for a country to exit the eurozone." [newspaper extract, from Gregoromichelaki to appear]

And the alleged common inferences with direct discourse are equally possible for such characterisations:

(58) In a reply to publications in the German newspapers Mario Draghi stated yesterday: "There is no practice or methodology for a country to exit the eurozone." [newspaper extract, from Gregoromichelaki to appear]

(59) Mario Draghi clarified that "there is no practice or methodology for a country to exit the eurozone." [from Gregoromichelaki to appear]

(60) Mario Draghi offered a clarification of his previous statements.

Such alleged "entailments" are not qualitatively different from the ones offered by G&C in (32)-(34). However, they cannot be explained as arising from a range of fixed speech-act specifications encoded in the grammar, which is what provides the explanation of (32)-(34) in the G&C model. If there is a mechanism for deriving the inferences in (58)-(60) pragmatically, it can also be used to derive the inferences in (32)-(34) as long as such pragmatically inferred contents can interact with grammatical specifications at an appropriate level.

On the other hand, the alleged inviolable restrictions posed for indirect reporting in the G&C account and others, regarding the interpretation of indexicals, do not hold for mixed quotation, a structurally similar construction as indirect quotation proper. So, for example, in a mixed quotation, a first person indexical need not refer to the speaker performing the utterance act but, instead, to the subject of the framing verb (Geurts and Maier 2005; Cumming 2005; Anand and Nevins 2003):

(61) Bill Watterson said that reality "continues to ruin my life". [from Maier 2014]

Additionally, *wh*-extraction is possible out of mixed quotation environments, which places mixed-quotation on a par with indirect discourse proper and indicates that quotation marks are not in any way "syntactic opacity" indicators (cf. Schlenker 2011 see e.g. (4) earlier), any extant constraints have to be sought elsewhere:

(62) Who did Mary say that she would "never misunderestimate ever again"? [from Maier 2014]

Maier (2014) claims, nevertheless, that certain features of the quoted original in mixed quotation have to be adjusted obligatorily to fit the new quoting environment. For example, he claims (citing Chung-chieh 2011) that grammatical gender agreement appearing in a quoted phrase in gender-determining languages has to be adjusted to fit its new environment. In our view of the data, there is no such strict requirement. There are examples like the following where this alleged restriction does not hold:

(63) *Ta koritsia tis Lenas ine poli psagmenes [Greek] [from Gregoromichelaki to appear]
The girls$_{\text{NEUT}}$ of Lena are very sophisticated$_{\text{FEM}}$
Lena's girls are very sophisticated

(64) I Maria ipe oti ta koritsia tis Lenas ine poli "psagmenes" [Greek] [from Gregoromichelaki to appear]
Mary said that her girls$_{\text{NEUT}}$ of Lena are very "sophisticated$_{\text{FEM}}$"
Mary said that Lena's girls are very sophisticated

We conclude that these intermediate phenomena, free (in)direct quotation, hybrid, and mixed quotation show that there is no strict distinction between direct and indirect reporting so that there is no need for distinct constructions to be defined

for each to account for their alleged distinct properties. Any such formalisation will prevent the whole range of phenomena from being captured. Speakers/writers can switch the mode of presentation of their utterance, indicate who takes responsibility for its content and form, or demonstrate some of its properties freely at any sub- or supra-sentential level. This argues against a model of NL-grammar that ignores the psycholinguistically established incrementality of processing, as well as the fact that grammatical semantic/syntactic constraints are not qualitatively different from pragmatic processing, and, therefore, cannot be segregated in a distinct abstract model. This is shown most clearly by the fact that contents provided by NL-sentences can compose with a variety of demonstrating events, like gestures, noises, laughs etc.:

(65) The car engine went [brmbrm], and we were off. [from Clark and Gerrig 1990]
(66) The boy who had scratched her Rolls Royce went [rude gesture with hand] and ran away. [from Recanati 2010]

To capture such phenomena, we now turn to a grammar formalism that takes into account the fact that language is primarily a form of action, produced and interpreted in context in a time-linear manner. We aim to show that the problematic for other formalisms data mentioned above find natural explanations from such a perspective.

6 Dynamic Syntax

In distinguishing between open and closed quotation (see earlier Sect. 5), Recanati (2010) makes an alleged important distinction: open quotations are primarily "pictorial" involving

> "the meaning of the speaker's act of ostensive display. That meaning is pragmatic: it is the meaning of an act performed by the speaker, rather than the semantic content of an expression uttered by the speaker" (Recanati 2010: 271).

Closed quotations in contrast, according to Recanati, carry additional referential meaning due to their integration in the linguistic system. In our view, this distinction reflects the standard conception of NL-analysis as requiring a grammar on the one hand and a separate component of pragmatic inference on the other (see also Capone 2013). In contrast, a more radical alternative concerning the status of the syntax/semantics components of the grammar and their integration with pragmatics is proposed by Dynamic Syntax (DS, Kempson et al. 2001; Cann et al. 2005; Gregoromichelaki and Kempson 2013).

DS models the act(ions) interlocutors engage in during the derivation of both meaning and forms. So all levels of NL analysis are reconceptualised as actions performed and assigned meaning in context. So DS can be seen as a psycholinguistically-inspired action-based formalism that specifies the 'know-how' that is employed in linguistic processing, in contrast to standard formalisms which codify (specifically linguistic) propositional knowledge of rules and representations. Regarding levels of analysis, DS eschews a string-syntactic level of explanation and

implements the assumption that grammatical constraints are all defined procedurally in terms of the progressive development of representations of content ('information states'), with partial interpretations emerging step-by-step during social interaction on a more or less word-by-word basis. In the view we sketch here, this is a variant which combines Dynamic Syntax with the Type Theory with Records framework (TTR, Cooper 2005, 2012) (DS-TTR), which captures directly the fine-grained dynamics of dialogue, as well as the potential for underspecification and enrichment (Purver et al. 2010; Eshghi et al. 2015).

DS-TTR is formulated as a system which crucially involves:

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

We will not go into the details of the formalism and the computations here;[2] for our purposes it suffices to look more closely at how this perspective, when applied to dialogue modelling and quotation devices, sheds new light on several puzzles: the phenomenon of split utterances seen earlier in (9), which we take up in Sect. 6.2; and how the mechanisms applied there, in combination with some of the tools provided by the G&C account, provide the means to model the continuity of direct and indirect discourse as we will see in Sect. 6.3. Since both dialogue phenomena and reporting devices are using the same grammatical resources they are predicted to interact. We show that DS-TTR is well-suited to account for such interactions.

6.1 Incrementality/Predictivity and Radical Contextual-Dependency in the Grammar

Instead of deriving sentence structures and propositional meanings, the DS grammar models the word-by-word processing of NL structures in context. For NL use in conversation this is a crucial explanatory factor since many of its features rely on such incremental production and comprehension. For example, the frequent occurrence of clarification requests in conversation (Ginzburg 2012 *inter alia*) shows that utterances can be processed and understood partially without having to map a sentential structure to a full proposition (contra in fact to Ginzburg 2012). Moreover, the process of grounding, modelled by Ginzburg (2012) (see Sect. 4.2.1 earlier) relies on the positioning of items like inserts, repairs, hesitation markers etc., a positioning which is not arbitrary but systematically interacts with grammatical categories and derivations at a sub-sentential level (see e.g. Clark and Fox Tree 2002 *inter alia*). In consequence, addressees display their comprehension and

[2]We cite throughout the publications where the relevant formal details can be found, and also see Gregoromichelaki (to appear).

assessments of the speaker's contribution subsententially as the utterance unfolds through *back-channel* contributions like *yeah*, *mhm*, etc. (Allen et al. 2001). And speakers shape and modify their utterance according to such verbal and non-verbal responses they receive from hearers as their turn unfolds (Goodwin 1981). Hence the grammar must be equipped to deal with those in a timely and integrated manner, i.e., by providing syntactic licensing and semantic interpretation online. In addition, the turn-taking system (see, e.g., Sacks, Schegloff, and Jefferson 1974) seems to rely on the grammar, as it is based on the predictability of (potential) turn endings in order for the next speaker to time appropriately their (potential) entrance; in this respect, experimental evidence has shown that this predictability is grounded mostly on syntactic recognition rather than prosodic cues etc. (De Ruiter et al. 2006). Therefore, the DS model assumes a tight interlinking of NL perception/production which imposes top-down predictive processes at all stages so that coordination among participants is the outcome of the fact that the grammar consists of a set of licensed actions that both speakers and hearers have to perform in synchrony. These actions perform step-by-step a mapping from phonological strings to semantic representations or vice-versa.

In DS-TTR, the semantic contents derived by processing linguistic strings are represented as trees inhabited by record types (see earlier Sects. 4.1 and 4.1.2). The nodes of these semantic trees are annotated by terms in a typed lambda calculus,[3] with mother-daughter node relations corresponding to semantic predicate-argument structure. For example, the CONTENT field associated eventually with the string *John left* will be the functional application of the lambda term $\lambda x.Leave'x$, inhabiting the function daughter, to the conceptual representation derived by processing the name *John*, which for simplicity we annotate here *John'*.[4] Following Gregoromichelaki (2006), we also assume here that the formula derived, *Leave'(John')*, also includes the event/situation referred to as well as the world/circumstance of evaluation (Recanati 2004). We also assume that each CONTENT field derived at each subnode of the tree includes independently shiftable world/event parameters to account for well-known cases of differentiation among the parameters of evaluation for various predicates in a sentence:

(67) The fugitives are now in jail [from Enç 1986]

In addition, in order to deal with the interpretation of indexicals like *I*, *you*, *now*, etc., contextual parameters are recorded in a structured *CONTEXT* field[5] on which the *CONTENT* field depends. The CONTEXT field records the occurrence of each

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

[4]Two analyses for names currently co-exist in DS: (a) as constants resulting from the contextual enrichment of metavariables introduced by names, and (b) as iota-terms. We remain agnostic on this as it does not affect the issues we discuss here.

[5]The differentiation CONTEXT vs. CONTENT fields is for convenience of display only, it does not signify any substantial claim regarding any qualitative differentiation among the parameters handled.

word-utterance event (utterance action) including the agent (utterer, which can be distinct by agent taking responsibility for the action), the addressee (which can be distinct from all the present participants), time/location of the action (following the specification of *micro conversational events* in Poesio and Rieser 2010, see earlier Sect. 4.1.1), and the world parameter of the context. Concatenation of such subevents produces cumulative utterance events in parallel with the phase of functional application. For our purposes here, we note that there can be additional world and event parameters in the CONTENT field, freely introduced or via the actions of linguistic operators, with accessibility relations represented as TTR-dependencies among CONTENT and CONTEXT fields (to deal with phenomena where shift of evaluation occurs, e.g. conditionals, see Gregoromichelaki 2006). In the CONTEXT parameters, following Ginzburg and Cooper (2014), we now add an NL-use parameter, indicated as a metavariable Γ_U, to represent the characterisation of the utterance as an event/action conforming to the types of action specified in this resource.

The grammar operates by modelling word-utterance events as conditional actions, in effect *characters* defined in procedural terms, that check for the existence of contextual/semantic/structural specifications on the information state and, accordingly, execute a macro of sub-actions, extending the tree-representations, or aborting in case the conditions of use of the word are not satisfied in the current linguistic / non-linguistic context. There is also a set of conditional *computational actions* that can apply without the parsing/production of lexical types if the specified conditions apply in the current state of the information-state. Such actions, either predictively prepare the ground for the execution of further lexical actions, or perform housekeeping functions like composing contents via functional application of functor nodes to argument nodes and concatenating the sequential subevents to a cumulative event of utterance.

As in DRT (Kamp 1981; Kamp and Reyle 1993) and related frameworks (see also Jaszczolt 2005; Jaszczolt et al, this volume), semantic, truth-conditional evaluation applies solely to these contextually-enriched representations, hence no semantic content is ever assigned to structures inhabited by strings of words (sentences). However, unlike all these other models, truth-conditional evaluation applies incrementally, as each word is processed. The other distinguishing feature of DS, as compared to DRT, is that this process of progressive building of semantically transparent structures is the only notion of "syntax" admitted, in that there is no intermediate level of syntactic structuring where the string of words is assigned hierarchically organised constituency as either phrases or clauses. Such constituency is considered in DS as epiphenomenal on the function-argument semantic relations as typified in the lambda-calculus analyses of NL meanings. In consequence, all syntactic dependencies have been reformulated in procedural terms, including, in particular, the classical evidence for denying the direct correspondence between NL-structure and semantic content that led to accounts via transformations (long-distance dependencies, binding, quantification etc., see e.g. Kempson et al. 2001; Cann et al. 2005; Gregoromichelaki 2006, 2011, 2013a).

This model directly provides the mechanisms for accounting for split utterances and fragmentary discourse in dialogue (see (9) earlier), since both speaker and addressee perform the processing steps incrementally, guided not solely by the NL string, but also driven by *predictions* ('goals'). These goals are imposed by either the procedures associated with NL elements ('lexical actions') or are system-generated as general top-down computational goals to be achieved in the next steps. Simplifying for presentation purposes, for example, in English, with its characteristic SVO structure, a general computational goal will ensure that parsing/production starts with the expectation of a subject first, followed by a predicate afterwards. Subsequently the lexical entries for transitive verbs will introduce not only the conceptual content associated with the word but also the prediction/expectation that an argument, the object, will follow immediately afterwards and the event that is taken as the witness of the type derived by processing the clause (see earlier Sect. 4.1.2). Likewise for all other regularities occurring in English or any other NL "syntactic" structuring. Thus, *parsing* in DS incorporates elements of *generation* (production) through the constant formulation of predictions for what will ensue next. On the other hand, production exploits the parsing mechanism in that licensing of the generation of each word relies in checking that the string so far produced can deliver a conceptual representation that accords with the (partial) conceptual structure the speaker attempts to verbalise. As a result, speaker and hearer roles involve mirroring of each other's actions (Gregoromichelaki 2013a, b; Pickering and Garrod 2012).

6.2 Split Utterances in Dynamic Syntax

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

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

Split utterances are then unproblematically processable and, in fact, a natural consequence of such a fine-grained bidirectional incremental system: As goals are constantly generated by the grammar, to be achieved symmetrically by both the parser and the producer, the addressee/parser can await for input from the speaker in order to fulfil these goals. However, according to the grammar, such goals are

also what activates the search of the lexicon ('lexical access') in production in order to recover a suitable NL word for the concept to be conveyed. As a result, a current hearer/parser who achieves a successful lexical retrieval before processing the anticipated NL input provided by the previous speaker can spontaneously become the producer and take over verbalising the continuation of the utterance instead (for detailed analyses see Eshghi et al. 2010, 2011; Gargett et al. 2008, 2009; Gregoromichelaki et al. 2011, 2013a, b; Kempson et al. 2011; Purver et al. 2006, 2009, 2011).

We will now see how these mechanisms which licence split-utterances in conversation interact with the reporting and metalinguistic phenomena.

6.3 Metalinguistic Devices in DS-TTR

Bonami and Goddard (2008), despite their otherwise significant contribution in providing a syntactic analysis for reporting constructions, characterise mixed and open quotation as "syntactically quite uninteresting" while admitting that they pose serious semantic problems. From their perspective, which aims to characterise sentential units, the noted parallel use-mention aspects and shifts in these structures can be ignored as data. However, from the current point of view, where syntax and semantics employ the same mechanisms, there is no independent level of syntactic characterisation, and the same syntactic mechanisms apply both to supra-sentential and subsentential licensing, things are different. Firstly, there is a requirement to address the modelling of the "semantic" issues mentioned by Bonami & Goddard, and, secondly, in fact, we will aim to show that there are significant interactions between the linguistic form and semantic interpretation of such structures that cannot be attributed to independent syntactic vs. semantic or, even pragmatic, mechanisms. For this purpose, the DS-TTR model, enriched with some of the resources offered by Ginzburg and Cooper (2014) (G&H, henceforth) accounts for a wide range of reported speech phenomena (for formal details see Gregoromichelaki to appear).

The lexical action for a framing verb (e.g. a *verbum dicendi* like *say*) can be assumed to uniformly combine with similar semantic objects in both direct and indirect reports, without imposing some specific type of syntactic complement. The only factor that accounts for the alleged "syntactic" differences between direct and indirect quotation (e.g. word order, or *wh*-extraction see earlier Sect. 2) is that the actions induced by such verbs, like other verbs in English (e.g. *eat*) can include "object-drop", which in the DS-TTR account is modelled by allowing such verbs to take as their complement a *metavariable*. As in the DS-TTR modelling of pronominal or elliptical anaphora resolution, such a metavariable has to be provided with a value from context. In direct quotation cases, the value for such a metavariable will be provided by the independent clause provided either anaphorically or cataphorically as an antecedent (see e.g. (69) below and (1)-(2) in Sect. 2 earlier).

(69) John shouted, "I talk better English than the both of youse!"

Such verbs, and various others, can also compose directly with non-linguistic actions, which is straightforwardly modelled in the DS-TTR formalism, as there is no qualitative distinction between "grammatical" and other actions:

(70) The car engine went [brmbrm], and we were off. [from Clark and Gerrig 1990]

(71) The boy who had scratched her Rolls Royce went [rude gesture with hand] and ran away. [from Recanati 2010]

This assumption allows us to capture the continuity of direct/indirect discourse as it appears in mixed quotation and free (in)direct discourse structures. The only differences among them occur in the specification of the CONTEXT field, which, we assume, is subject to pragmatic enrichment (Recanati 2010), at a subsentential level, so that the truth-conditional content derived is always directly affected.

As the DS-TTR grammar is articulated in terms of actions, we can postulate that the added properties that characterise direct discourse are the result of focusing the hearer's attention to the actions used by another speaker, whether at the level of types or particular tokens, which, we assume, can sometimes be indicated by the quotation marks. Essentially we agree with Maier (2014) that direct discourse and mixed quotation are the same phenomenon, however, in line with G&C, we don't think that the grammar needs to implement this insight by employing special devices. Since we do not posit a separate level of syntactic analysis for the string of words, only the semantic-conceptual representation derived by processing the string, there is no issue arising here in terms of characterising a distinct syntactic category for indirect, direct, and mixed quotation complements in contrast to any other grammatical analysis of quotation (and, in fact, contra Recanati 2000, 2010). In consequence, unlike G&C, we do not employ specific constructions to deal with separate quotational phenomena, so that the present account extends naturally to mixed quotation and free (in)direct discourse.

To distinguish the properties of what we assume are variable ways of processing, in line with Recanati (2010), we analyse standard uniform indirect reports as cases where the CONTEXT field remains stable throughout the utterance of both the reporting section of the sentence and the reported-event part. As a consequence, indexical elements receive their interpretations from the context established by the current utterance event U_0. However, as a consequence of the lexical action introduced by the framing verb, a new world parameter W_1 is introduced for the report to express the reportee's view. The contextual and world parameters can be shifted independently of each other, and the possibility of shifting world and context parameters incrementally as the utterance develops models the otherwise puzzling cooccurrences of transposed and untransposed indexicals considered by Recanati (2000: Ch 15–16).

Note that in cases of shared indirect reports, indexicals will acquire values according to who currently assumes the relevant interlocutor roles (see also (18)-(20) earlier):

(72) A: So **you** say **you** will live
B: by **my** pen, yes
(73) A: Did **you** say to Nick that . . .
B: **you** hate **me**? Yes, I think it's true ('A hates B')
(74) A: Did **you** say to Nick that **you** . . .
B: hate **myself**? Yes, my psychoanalyst says so. ('B hates B')

As we said earlier in Sect. 6.1, the eventual representation derived, following standard DS-TTR procedures, composes the contents derived at the various subsentential stages, as well as recording the various concatenated u_0 subevents that resulted in a (perhaps joint) utterance-event U_0. Hence the interpretation derived eventually has the values of the indexicals as intended by the participants at each previous stage in that their "characters" (lexical actions in DS) have been applied subsententially to the then current context so that the eventual composition deals with contents only. The fact that there is no level of syntactic representation for the string of words makes utterances like (74) fully-licensed as joint utterances and provided with appropriate interpretations (the same for (18)-(20) seen earlier). Any other grammar that insists on an independent syntactic analysis of such strings (see e.g. Potts 2007; Maier 2014) will have trouble with such utterances as the string of words *Did you say to Nick that you hate myself* will have to be characterised as ungrammatical (and for (73) will derive the wrong interpretation).

Following G&C, in direct discourse, a new utterance event U_D is introduced, corresponding to the demonstration the speaker performs. As in G&C, this newly introduced event bears a contextually-determined similarity value (resemblance) to another, anaphorically-retrieved utterance event U_Q from which contextual parameters are copied, thus accounting for the corresponding change in the values of indexicals that can be stable across speakers and turns:

(75) Adam: Well. I can tell you what her view on that is. and that
Sherm: what.
Adam: is, .h **I**'m older, and therefore I'm in a worse competitive position, and **I** and **I**'ve really got to produce.
Sherm: but **I**'m smarter== (laughs) yeah. (said very softly)
Adam: and **I**'m going to.
Sherm: yeah. (said very softly) [from Grimshaw 1987]
(76) A: Did **you** say to Nick. . .
B: "**I** hate **you"**? Yes, why? ('B hates Nick')
(77) A: Did **you** say to Nick "**You** . . .
B: "hate **yourself"**? Yes, why? ('Nick hates Nick')

Given that linguistic and non-linguistic actions are not differentiated in this account, the fact that the demonstrating event offers anaphoric possibilities that can be exploited subsequently both supra-sententially, subsententially, and across turns is a natural prediction:

(78) "I talk better English than the both of youse!" shouted Charles, thereby convincing me that he didn't. [from Partee 1973]

(79) "Don't worry, my boss likes me! He'll give me a raise" said Mary, but given the economic climate I doubt that he can. [from Maier 2015]

(80) A: I talk better English than the both of youse!
B: You obviously don't [from Partee 1973]

Instead of assuming that the availability of such anaphoric resolutions is the result of presuppositional elements or implicatures (as in Maier 2014), here the grammar itself provides the resources for explaining the phenomena. The resolution of both ellipsis and pronominal anaphora in DS is assumed to involve the reuse of terms annotating CONTENT fields on treenodes or the rerunning of processing actions stored in the context (Kempson et al. 2012; Kempson et al. to appear). Since the demonstrating event is constituted by a set of such processing actions, and both the ensuing content and its processing actions are not segregated from the rest of the discourse representation, they are stored in the context and are available to be invoked for the resolution of anaphoric occurrences as in (78)-(80). For the same reason, as in G&C analysis, we can account for cases of "mixed predication" where both token and type aspects are addressed simultaneously:

(81) 'Was I snoring' was asked by Bill and is a frequently used interrogative clause.

(82) 'Am I snoring?' asked Bill, a sentence frequently uttered by men who don't think they snore. It is usually answered by 'You were before you woke up.'

But further than the G&C account, the present analysis extends to cases where a continuation of an utterance started with an initial speaker without any quotational "intentions" can become quotational, i.e., treated as a demonstrating event, and conversely, structures initiated without necessary provision of a quotation can be provided a quotational, echoing, complement:

(83) Jem: Mary, whatever it is you think you know you mustn't speak of it. Not if you want to stay safe.
Mary: says the horse-thief [Jamaica Inn BBC, Episode 1, 23:50'] [from Gregoromichelaki, to appear]

(84) Miriam: That is the nastiest, dirtiest thing anyone has ever done
Patience: says Black Peter's strumpet! What are you crying for?
[from Gregoromichelaki, to appear]

(85) Noel: What I'm saying is
Stacey: you are It!
Noel (ironically): Yeah . . . [from Gregoromichelaki, to appear]

The difference between this account and G&H is that a monolithic demonstration event U_D is not necessarily derived at once for the whole complement of the framing verb. Instead, there is the possibility at each subsentential stage for the speaker to switch in and out of the demonstration. This is what accounts for both cases of free indirect reports and mixed quotation.

In the cae of *free indirect discourse*, in addition to the introduction of a demonstrating event, there is also a shift in the world parameter of the context (Recanati 2000), for example, the event is taking place in a world according to somebody's beliefs:

(86) John is totally paranoid. Everybody spies on **him** or wants to kill **him**, including **his** own mother. [from Recanati 2000]

Since there is the possibility for independent shifting of world and context parameters and the eventual interpretation emerges via the concatenation of utterance subevents which can define their contexts independently of each other, corresponding to the sequential shifting in and out of demonstrations the speaker performs, there is no problem with having to coordinate the world and context shifts. This account gives results similar to Meier (2015) but without using ad hoc devices like the "unquotation" mechanism. The results just follow from the incremental contextual licensing of structures and interpretation that constitute ab initio the basis of the model. And, unlike other grammatical analyses, e.g. G&C, Sharvit (2008), since there is no independent level of syntactic analysis for the sentence, we do not have to license a complete sentential complement that has to be internally consistent as to indirect/direct report features and contextual parameters (remember we compose contents not characters). So *free direct discourse* (see (36) in Sect. 5) is simply a case where the context parameters are also shifted along with the world parameter.

In the cases of *mixed quotation* (seen earlier in (41) in Sect. 5) we assume that there is no "verbatim requirement" (cf. Meier 2014), so no such difference with indirect discourse ensues, since, as Recanati (2010) has pointed out, the context might make it evident that the words of somebody else rather than the subject of the framing verb are being echoed, or, we would add, that nobody has in fact uttered those words (we do not consider 'scare quoting' as a separate phenomenon):

(87) Alice said that Clinton is 'smooth', as you would put it. Of course that's not the word SHE used. [from Recanati 2010]

(88) These are not 'I really should' radishes . . . . [from Clark and Gerrig 1990: ex. 5b]

(89) Dutch is a "that I him have helped" language. [from Abbott 2005, attributed to Philippe de Brabanter]

Non-constituent mixed quotation does not present a fundamental problem for this account either, since, by definition, the grammar licenses and interprets incrementally word strings, without relying on what other grammars characterise as "syntactic constituents" either subsententially or supra-sententially:

(90) She allowed as how her dog ate "strange things, when left to its own devices". [from Abbott 2005]

(91) Pascal suspected that the mercury was really supported by the "weight and pressure of the air, because I consider them only as a particular case of a universal principle concerning the equilibriums of fluids." [from Maier 2008]

(92) Also, he categorically stated that "there is no legal way of temporal extension of the Greek debt without this being regarded as a credit event. Therefore there is no way that it will be allowed to happen such a credit event in Greece because it would create negative impact on the whole system." [from Gregoromichelaki, to appear]

But we can go even further than that to account for data that are completely out of reach for other grammars. Given its psycholinguistically-inspired nature, the DS-TTR account models the various alternative options arising during the processing

of ambiguous strings. Even options less probabilistically favoured and, hence, not currently pursued, are stored temporarily in the context model (see e.g. Hough 2015) in order to be employed for, e.g., the functioning of repair processes, like corrections, in dialogue. This independently needed modelling allows us here to capture the variable semantic-"constituency" ambiguity of some mixed quotation strings and how they can be exploited by interlocutors, for example, in puns and jokes, as pointed out by Maier (2014):

(93) The menu says that this restaurant serves "[breakfast] [at any time]" . . . [so I ordered [French toast during the Renaissance]]. [Steven Wright joke, mentioned in Maier 2014]

Even though, for reasons of space, we cannot go into all the details here, all the "peculiarities" of mixed quotation presented in Maier (2014) and others, e.g. quantifier-"raising" blocking etc. are eliminated in DS-TTR, since there is no level of syntactic constituency or any independent syntactic categories assumed (see Gregoromichelaki to appear for further formal details). One might wonder at this stage as to whether exactly this assumption will prevent us from dealing with cases of *pure quotation*, i.e. cases where metalinguistic mention is made to folk-linguistic categories, like "sentences", "phrases", "words" etc. As argued in Gregoromichelaki (to appear), we do not believe that we are at a disadvantage here. Unlike G&C, we do not assume that folk-linguistic characterisations and reifications coincide with what the psycholinguistic account provides. Instead such characterisations, like speech-act characterisations (see Allen, this volume), can be freely pragmatically generated in an ad hoc manner that suits the interlocutors' purposes and there is no requirement for systematicity and consistency at least for their use in informal everyday discourse. For this purpose, our modification of the parameter in the CONTEXT field, borrowed from Recanati (2010) and G&C (see earlier Sect. 4.2.3), allows for the anaphoric resolution of an NL-use parameter Γ that can be exploited in determining appropriate referents for such purely metalinguistic uses. An essential difference with the G&C account though is that the quoting event in such cases does not project a particular phrase or construction so that it can combine with the rest of the sentence. Instead it can appear directly as the argument of an appropriate predication since utterance event act(ion) and the grammatical actions induced by linguistic input mesh together without further ado as we also saw in cases like (70)-(71) in Sect. 6.3 earlier.

7 Conclusion

We have seen now that in taking a psycholinguistically-realistic view of grammar, one that relies on the incrementality/predictivity and contextual dependence of NL-processing, we can accommodate not only various recalcitrant for others dialogue phenomena, like fragmental and split utterances, but also the various uses words are put to in echoing others' speech and thought or even referring to the function of

the NL-system itself. We have also argued that, since all these phenomena employ the same mechanisms (conceptualised as actions) for their manifestation, we would expect, and we do find, significant interactions among them. We have sketched a model where such linguistic, metalinguistic and non-linguistic factors mesh together in deriving the coordination of action that characterises human interaction.

References

Abbott, B. (2005). Some notes on quotation. *Belgian Journal of Linguistics, 17*, 13–26.

Allen, J., George, F., & Amanda, S. (2001). *An architecture for more realistic conversational systems*. Proceedings of the 2001 international conference on Intelligent User Interfaces (IUI), January 2001.

Anand, P., & Nevins, A. (2003). Shifty operators in changing contexts. In *Proceedings of semantics and linguistic theory* (pp. 20–37). Ithaca: CLC Publications.

Antaki, C., Díaz, F., & Collins, A. F. (1996). Keeping your footing: Conversational completion in three-part sequences. *Journal of Pragmatics, 25*(2), 151–171.

Arundale, R. B., & Good, D. (2002). Boundaries and sequences in studying conversation. In F. Anita Fetzer & M. Christiane (Eds.), *Rethinking sequentiality. Linguistics meets conversational interaction* (pp. 121–150). Amsterdam: John Benjamins.

Barwise, J., & Perry, J. (1983). *Situations and attitudes*. Cambridge: MIT Press.

Blakemore, D. (1987). *Semantic constraints on relevance*. New York: Wiley-Blackwell.

Bonami, O., & Godard, D. (2008). On the syntax of direct quotation in French. In S. Müller (Ed.), *Proceedings of the HPSG08 conference* (pp. 358–377). Stanford: CSLI Publications.

Cann, R., Kempson, R., & Marten, L. (2005). *The dynamics of language*. Oxford: Elsevier.

Capone, A. (2013). The pragmatics of quotation, explicatures and modularity of mind. *Pragmatics and Society, 4*(3), 259–284.

Cappelen, H., & Lepore, E. (1997). Varieties of quotation. *Mind, 106*, 429–450.

Chung-Chieh, S. (2011). The character of quotation. Linguistics and Philosophy, 33(5), 417–443. http://dx.doi.org/10.1007/s10988-011-9085-6..

Clark, H. H. (1996). *Using language*. New York: Cambridge University Press.

Clark, H. H., & Fox Tree, J. E. (2002). Using uh and um in spontaneous speech. *Cognition, 84*, 73–111.

Clark, H. H., & Gerrig, R. J. (1990). Quotations as demonstrations. *Language, 66*, 764–805.

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

Cooper, R. (2012). Type theory and semantics in flux. In R. Kempson, N. Asher, & T. Fernando (Eds.), *Philosophy of linguistics* (Handbook of the philosophy of science, Vol. 14). Amsterdam: Elsevier.

Crystal, D. (2013). http://www.davidcrystal.community.librios.com/?id=2914. Accessed 13 Jan 2015.

Cumming, S. (2005). Two accounts of indexicals in mixed quotation. *Belgian Journal of Linguistics, 17*, 77–88.

Davidson, D. (1968). On saying that. *Synthese, 19*, 130–146.

Davidson, D. (1984). Quotation. In D. Davidson (Ed.), *Inquiries into truth and interpretation* (pp. 79–92). Oxford: Clarendon.

de Brabanter, P. (2010). The semantics and pragmatics of hybrid quotations. *Language and Linguistics Compass, 4*(2), 107–120. http://dx.doi.org/10.1111/j.1749-818X.2009.00185.x.

Enç, M. (1986). Towards a referential analysis of temporal expressions. *Linguistics and Philosophy, 9*, 405–426.

Eshghi, A., Healey, P. G. T., Purver, M., Howes, C., Gregoromichelaki, E., & Kempson, R. (2010, September). Incremental turn processing in dialogue. In *Proceedings of the 16th annual conference on Architectures and Mechanisms for Language Processing (AmLAP).* New York.

Eshghi, A., Purver, M., & Hough, J. (2011). *Dylan: Parser for dynamic syntax*. Technical report, Queen Mary University of London.

Eshghi, A., Howes, C., Gregoromichelaki, E., Hough, J., & Purver, M. (2015). Feedback in conversation as incremental semantic update. In *Proceedings of the 11th international conference on Computational Semantics (IWCS 2015)* (pp. 261–271). London: Queen Mary University of London.

Gargett, A., Gregoromichelaki, E., Howes, C., & Sato, Y. (2008). *Dialogue-grammar correspondence in dynamic Syntax*. Proceedings of the 12th SemDial (LonDial).

Gargett, A., Gregoromichelaki, E., Kempson, R., Purver, M., & Sato, Y. (2009). Grammar resources for modelling dialogue dynamically. *Journal of Cognitive Neurodynamics, 3*(4), 347–363.

Geach, P. (1957). *Mental acts*. London: Routledge Kegan Paul.

Geurts, B., & Maier, E. (2005). Quotation in context. *Belgian Journal of Linguistics, 17*, 109–128.

Ginzburg, J. (2012). *The Interactive stance: Meaning for conversation*. Oxford: Oxford University Press.

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

Ginzburg, J., & Cooper, R. (2014). Quotation via dialogical interaction. *Journal of Logic, Language and Information, 23*, 287–311.

Goffman, E. (1979). Footing. *Semiotica, 251–2*, 1–30.

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

Gregoromichelaki, E. (to appear). Quotation in dialogue. In P. Saka & M. Johnson (Eds.), *The pragmatics of quotation*. Springer

Gregoromichelaki, E. (2006). *Conditionals in dynamic syntax*. PhD thesis. University of London.

Gregoromichelaki, E. (2011). Conditionals in dynamic syntax. In R. Kempson, E. Gregoromichelaki, & C. Howes (Eds.), *The dynamics of lexical interfaces*. Stanford: CSLI Publications.

Gregoromichelaki, E. (2013a). Clitic left dislocation and clitic doubling: A dynamic perspective on left-right asymmetries in Greek. In G. Webelhuth, M. Sailer, & H. Walker (Eds.), *Rightward movement in a comparative perspective*. Amsterdam: John Benjamins.

Gregoromichelaki, E. (2013b). Grammar as action in language and music. In M. Orwin, C. Howes, & R. Kempson (Eds.), *Language, music and interaction*. London: College Publications.

Gregoromichelaki, E., & Kempson, R. (2013). The role of intentions in dialogue processing. In A. Capone, F. Lo Piparo, & M. Carapezza (Eds.), *Perspectives on linguistic pragmatics* (Perspectives in pragmatics, philosophy & psychology, Vol. 2). New York: Springer.

Gregoromichelaki, E., & Kempson, R. (2015). Joint utterances and the (split-)turn taking puzzle. In A. Capone & J. Mey (Eds.), *Interdisciplinary studies in pragmatics, culture and society* (pp. 703–744). Switzerland: Springer.

Gregoromichelaki, E., Kempson, R., Purver, M., Mills, G. J., Cann, R., Meyer-Viol, W., & Healey, P. G. T. (2011). Incrementality and intention-recognition in utterance processing. *Dialogue and Discourse, 2*(1), 199–233. special issue on Incremental Processing in Dialogue.

Gregoromichelaki, E., Kempson, R., Howes, C., & Eshghi, A. (2013a). On making syntax dynamic: The challenge of compound utterances and the architecture of the grammar. In W. Ipke, J. de Ruiter, J. Petra, & K. Stefan (Eds.), *Alignment in communication: Towards a new theory of communication* (Advances in interaction studies). Amsterdam: John Benjamins.

Gregoromichelaki, E., Cann, R., & Kempson, R. (2013b). Coordination in dialogue: Subsentential speech and its implications. In L. Goldstein (Ed.), *Brevity*. Oxford: Oxford University Press.

Grimshaw, A. D. (1987). Finishing other's talk: some structural and pragmatic features of completion offers. In R. Steele & T. Threadgold (Eds.), *Language topics, essays in honor of Michael Halliday* (pp. 213–35). Amsterdam: John Benjamins.

Groenendijk, J., & Stokhof, M. (1991). Dynamic predicate logic. *Linguistics and Philosophy, 14*(1), 39–100.

Henetz, T. & Clark, H. H. (2011) Managing delays in speaking. Paper presented at the IPrA 1413 meeting, July 2011, Manchester.

Hough, J. (2015). *Modelling incremental self-repair processing in dialogue*. PhD thesis. Queen Mary University of London.

Huddleston, R., & Pullum, G. (2002). *The Cambridge grammar of the English language*. Cambridge: Cambridge University Press.

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

Kamp, H. (1981). A theory of truth and semantic interpretation. In J. Groenendijk, T. Janssen, & M. Stokhof (Eds.), *Formal methods in the study of language* (pp. 277–322). Amsterdam: Amsterdam Center.

Kamp, H., & Reyle, U. (1993). *From discourse to logic. Introduction to model-theoretic semantics of natural language, formal logic and discourse representation theory*. Dordrecht: Kluwer.

Kaplan, D. (1989). Demonstratives. In A. Joseph, P. John, & W. Howard (Eds.), *Themes from Kaplan* (pp. 481–563). Oxford: Oxford University Press.

Kempson, R., Gregoromichelaki, E., Eshghi, A., & Hough, J. (to appear). Ellipsis in dynamic syntax. In van Craenenbroeck, J., & Temmerman, T. (eds.). *The Oxford handbook of ellipsis*. Oxford University Press.

Kempson, R., Meyer-Viol, W., & Gabbay, D. (2001). *Dynamic syntax: The flow of langage understanding*. Oxford: Blackwell.

Kempson, R., Gregoromichelaki, E., Meyer-Viol, W., Purver, M., White, G., & Cann, R. (2011). Natural-language syntax as procedures for interpretation: the dynamics of ellipsis construal. In *Proceedings of the PRELUDE workshop on games, Dialogues and Interaction. Lecture Notes in Computer Science*. Berlin: Springer.

Kempson, R., Gregoromichelaki, E., & Chatzikyriakidis, S. (2012) Joint utterances in Greek: their implications for linguistic modelling. In *Proceedings of 33rd annual linguistics meeting Syntactic Theories and the Syntax of Greek* (pp. 26–27). Thessaloniki.

Köder, F., & Maier, E. (2015). Children mix direct and indirect speech: Evidence from pronoun comprehension. (prefinal version) Online First. *Journal of Child Language*. doi:10.1017/S0305000915000318.

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

Larsson, S., & Traum, D. R. (2000). Information state and dialogue management in the TRINDI dialogue move engine toolkit. *Natural Language Engineering, 6*(3–4), 323–340.

Larsson, S. (2011) *The TTR perceptron: Dynamic perceptual meanings and semantic coordination*. In Proceedings of the 15th Workshop on the Semantics and Pragmatics of Dialogue (SemDial 2011 – Los Angelogue) (pp. 140–148), Los Angeles.

Levinson, S. (2000). *Presumptive meanings*. Cambridge: MIT Press.

Maier, E. (2007). Mixed quotation: Between use and mention. In *Proceedings of LENLS 2007*, Japan: Miyazaki.

Maier, E. (2008). Breaking quotations. In S. Ken, I. Akihiro, N. Katashi, & K. Takahiro (Eds.), *New frontiers in artificial intelligence* (Lecture notes in computer science, Vol. 4914, pp. 187–200). Berlin/Heidelberg: Springer. http://dx.doi.org/10.1007/978-3-540-78197-4_18.

Maier, E. (2014). *Mixed quotation*. Ms. University of Groningen [survey article for *Blackwell Companion to Semantics*], Groningen.

Maier, E. (2015). Quotation and unquotation in free indirect discourse. *Mind & Language, 30*(3), 345–373. doi:10.1111/mila.12083.

McCloskey, J. (2006). Questions and questioning in a local English. In R. Zanuttini, H. Campos, E. Herburger, & P. H. Portner (Eds.), *Cross-linguistic research in syntax and semantics: Negation, tense and clausal architecture* (pp. 87–126). Washington: Georgetown University Press.

Partee, B. (1973) The syntax and semantics of quotation. In *A festschrift for Morris Halle* S.R. Anderson and P. Kiparsky (eds.), New York: Holt, Reinehart and Winston, (pp. 410–418).

Pickering, M., & Garrod, S. (2012). An integrated theory of language production and comprehension. *Behavioral and Brain Sciences., 36*, 329–347.
Poesio, M., & Rieser, H. (2010). Completions, coordination, and alignment in dialogue. *Dialogue and Discourse, 1*(1), 1–89.
Poesio, M., & Traum, D. R. (1997). Conversational actions and discourse situations. *Computational Intelligence, 13*(3).
Poesio, M., & Traum, D. (1998). *Towards an axiomatization of dialogue acts*. Proceedings of the Twente Workshop on the Formal Semantics and Pragmatics of Dialogues (13th Twente Workshop on Language Technology).
Potts, C. (2007). The dimensions of quotation. In C. Barker & P. Jacobson (Eds.), *Direct compositionality* (pp. 405–431). New York: Oxford University Press.
Purver, M., Cann, R., & Kempson, R. (2006). Grammars as parsers: Meeting the dialogue challenge. *Research on Language and Computation, 4*(2–3), 289–326.
Purver, M., Howes C., Gregoromichelaki, E., & Healey, P. G. T. (2009 September). Split utterances in dialogue: a corpus study. In *Proceedings of SigDial*, London.
Purver, M., Gregoromichelaki, E., Meyer-Viol, W., & Cann, R. (2010, June). Splitting the I's and Crossing the You's: Context, Speech Acts and Grammar. In *SemDial 2010 (PozDial)*, Poland: Poznan.
Purver, M., Arash E., & Julian, H. (2011, January). Incremental semantic construction in a dialogue system. In *9th international conference on Computational Semantics (IWCS)*, Oxford.
Quine, W. V. O. (1940). *Mathematical logic*. Boston: Harvard University Press.
Recanati, F. (2000). *Oratio recta, oratio obliqua*. Cambridge: MIT Press.
Recanati, F. (2001). Open quotation. *Mind, 110*, 637–687.
Recanati, F. (2004). *Literal meaning*. Cambridge: Cambridge University Press.
Recanati, F. (2010). *Truth-conditional pragmatics*. Oxford: Oxford University Press.
Ruiter, D., Jan-Peter, H. M., & Enfield, N. J. (2006). Projecting the end of a speakers turn: A cognitive cornerstone of conversation. *Language, 82*(3), 515–535.
Ruth, K., Gregoromichelaki, E., & Howes, C. (2011). *The dynamics of lexical interfaces*. Stanford: CSLI Publications.
Sacks, H., Schegloff, E. A., & Jefferson, G. (1974). A simplest systematics for the organization of turn-taking for conversation. *Language, 50*, 696–735.
Saka, P. (1998). Quotation and the use-mention distinction. *Mind, 107*(425), 113–135.
Saka, P. (2011). Quotation and conceptions of language. *Dialectica, 65*(2), 205–220.
Schegloff, E. A. (2007). *Sequence organization in interaction: A primer in conversation analysis*. Cambridge: Cambridge University Press.
Schlenker, P. (2011). Indexicality and De Se reports. In C. Maienborn, K. von Heusinger, & P. Portner (Eds.), *Semantics: An international handbook of natural language meaning* (Vol. 2, pp. 1561–1604). Berlin: Mouton de Gruyter.
Sharvit, Y. (2008). The puzzle of free indirect discourse. *Linguistics and Philosophy, 31*(3), 353–395. doi:10.1007/s10988-008-9039-9.
Sperber, D., & Wilson, D. (1995). *Relevance*. Oxford: Blackwell.
Stone, M. (2004). Intention, interpretation and the computational structure of language. *Cognitive Science, 28*(5), 781–809.
Tarski, A. (1993). The concept of truth in formalized languages. In A. Tarski (Ed.), *Logic, semantics, metamathematics* (2nd ed., pp. 152–278). Indianapolis: Hackett.
Washington, C. (1992). Quotation. *Journal of Philosophy, 89*, 582–605.
Wilson, D., & Sperber, D. (1988). Mood and the analysis of non-declarative sentences. In J. Dancy., J. Moravcsik., & C. Taylor (Eds.), *Human agency: Language, duty and value* (pp. 77–101). Stanford: Stanford University Press. Reprinted In A. Kasher (Ed.), (1998) *Pragmatics: Critical concepts* (Vol. II) (pp. 262–289). London: Routledge.