Quotation in Dialogue

Eleni Gregoromichelaki

Abstract Quotation is ubiquitous in natural language (NL). Recent grammars that take a dialogical view on the formal and semantic properties of NLs (Ginzburg, The interactive stance: meaning for conversation. Oxford University Press, Oxford, 2012; Gregoromichelaki et al. Dialog Discourse 2(1):199–233, 2011; Eshghi et al. Feedback in conversation as incremental semantic update. In Proceedings of the 11th International Conference on Computational Semantics (IWCS 2015), Queen Mary University of London, UK April 2015, 261–271, 2015) indicate that quotation mechanisms need to be integrated within the purview of standard grammatical frameworks since such mechanisms are crucially involved in metacommunicative conversational interaction. Accordingly, the account presented in Ginzburg and Cooper (J Logic Lang Inf 23(3):287–311, 2014, G&C) provides syntactic analyses, denotations, and pragmatic constraints for quotational constructions that make use of grammatical entities independently needed for the analysis of conversation. However, despite the great advances achieved by G&C, the construction-based grammar employed lacks essential integration of the psycholinguistically grounded observation that NL use relies crucially on incremental/predictive processing with context integration at each word-by-word processing stage. For this reason, certain data showing the grammatical continuum underpinning various quotational constructions as well as interactions between quotation mechanisms and conversational phenomena (*split-utterances*, Gregoromichelaki et al. Dialog Discourse 2(1):199–233, 2011) are not amenable to G&C's discrete constructional approach. Based on this inadequacy of even such a state-of-the-art, comprehensive model, this chapter argues that a satisfactory account of the function of quotational devices cannot be given within standard NL theories involving the division of labour between syntax and semantics/pragmatics. Instead, it adopts a dynamic, incremental perspective that takes joint action as the basis for the definition of the grammar as advocated

E. Gregoromichelaki (✉)
Heinrich Heine Universität Düsseldorf, Düsseldorf, Germany

King's College London, London, UK
e-mail: elenigregor@gmail.com

P. Saka, M. Johnson (eds.), *The Semantics and Pragmatics of Quotation*,
Perspectives in Pragmatics, Philosophy & Psychology 15,
https://doi.org/10.1007/978-3-319-68747-6_8

within Dynamic Syntax (DS, Kempson et al. Dynamic syntax: the flow of language understanding. Blackwell, Oxford, 2001) updated with the integration of some of G&C's proposed formal constructs (DS-TTR, Purver et al. Splitting the *I*'s and crossing the *you*'s: Context, speech acts and grammar. In Proceedings of SemDial 2010 (PozDial), Poznan, Poland, 2010; Eshghi et al. Feedback in conversation as incremental semantic update. In Proceedings of the 11th International Conference on Computational Semantics (IWCS 2015), Queen Mary University of London, UK, April 2015, 261–271, 2015).

1 Introduction

It has long been noted that quotation is ubiquitous in natural language (NL), either obliquely in the form of *dialogism* or *heteroglossia* (Bakhtin 1981; Morris 1997), or directly with more or less explicit indications in conversation and written texts. It is puzzling then that both NL formal grammars and philosophical accounts (e.g. Davidson 1979) have assumed that quotation constitutes some sort of extraordinary or abnormal use, or that elements between (possibly implicit) quotation marks need not be generated by standard syntax or addressed by the semantics. Since Partee (1973), various counterexamples to this claim have been presented, for example cases where the formal or semantic properties of the quotation are needed to construct antecedents for anaphora/ellipsis:

(1) We had the "crap sticks", according to the translation on the menu, but **they** were actually delicious.
(2) She said "I've hiked in Grete" and I confirmed she **did**. Of course **it**'s called Crete.
(3) George says Tony is his "bestest friend" and indeed he **is**. (adapted from Geurts and Maier 2005)
(4) "I talk better English than the both of youse!" shouted Charles, thereby convincing me that he **didn't** (Partee 1973)

Other counterevidence too (examined later) favours the conclusion that quotation has to be included in the syntactic analysis and grammaticality definitions of any grammar (see also Postal 2004).

Recent grammars that take a dialogical view of the formal and semantic properties of NLs (Ginzburg 2012, Gregoromichelaki et al. 2011, Eshghi et al. 2015) provide additional arguments for integrating quotation within the purview of standard grammatical mechanisms. Quotation mechanisms are crucially involved in cases of metacommunicative exchanges in dialogue. One such case is the phenomenon of *echo questions*, for example *clarification requests,* whose content includes mentioning some previously uttered token (Noh 1998, Ginzburg 2012, Eshghi et al. 2015):

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

Currently a number of accounts have been proposed regarding the grammatical integration of quotational expressions. However, such accounts introduce ad hoc entities and rules in order to shift the usual contents attributed to NL elements and otherwise ignore the interaction of the whole grammatical apparatus, syntax, phonology, semantics, and, sometimes, even pragmatics in explaining the totality of the phenomenon. In contrast, the account of quotation presented in Ginzburg and Cooper (2014)—henceforth *G&C*—provides syntactic analyses, denotations, and pragmatic constraints for quotational constructions that utilise independently needed and antecedently established grammatical entities. In this chapter, the G&C insights are adopted to a large extent but within a distinct grammatical framework that combines Dynamic Syntax (e.g. Kempson et al. 2001; Cann et al. 2005) with Type Theory with Records (e.g. Cooper 2012; Ginzburg 2012). This combination, DS-TTR (Purver et al. 2010; Gregoromichelaki 2013b; Eshghi et al. 2015; Hough 2015), takes into account the fact that NL use crucially displays incrementality and predictivity in its processing with contextual integration at each word-by-word processing stage.[1] In DS-TTR, the incrementality and predictivity of linguistic processing are considered essential enough to be included as part of the design of the grammar formalism. As a consequence, certain quotation data that are not amenable to appropriate treatment within the G&C model become naturally accounted for under the DS-TTR modelling. For example, quotation can appear subsententially, and discontinuously at any point, which means that contextual parameters regarding the discourse situation and semantic evaluation variables need to be able to shift incrementally[2]:

(6) "Cities," he said, "are a very high priority." [FrameNet]

(7) Wright won't disclose how much the Nike deal is worth, saying only that "they treat me well". (De Brabanter 2010a, from *The Face*, September 93: 55)

(8) A doctor tells him [Gustave Flaubert] he is like a "vieille femme hystérique"; he agrees. (De Brabanter 2010a, from *TLS Online*, 18 December 1998)

(9) Alice said that life is "difficult to understand". (Cappelen and Lepore 1997)

(10) Mary felt relieved. If Peter came tomorrow, she would be saved. (free indirect speech; Recanati 2000)

Additionally, quotation is one of the environments where the phenomenon of *split-utterances* (Gregoromichelaki et al. 2011) is frequently observed due to the assumed projectibility of the upcoming continuation (Lerner 1991) and, in my view, the

[1] *Incrementality* here refers to the psycholinguistic notion of a stimulus' multi-level analysis piece-by-piece as it is temporally encountered (see e.g. Marslen-Wilson 1973; Steedman 1992; Tanenhaus et al. 1995; Chater et al. 1995).

[2] Elements in the data that constitute the main focus of a claim appear shaded in the examples. Bold font indicates additional highlighting.

opportunity for co-constructing a vivid unified perspective on some other (actual or imaginary) speech/thought event:

(11) Clinician: So I watch … this person being killed and then I go to bed and I'm you know lying there going, "well."
Patient: "Did I hear something?" (Duff et al. 2007)

(12) Patient: And it's like well … you know "I just-"
Clinician: "I don't want to." (Duff et al. 2007)

(13) Joan: she wz wise she'd pick up the phone en say, …
Linda: I'm comin over, (Lerner 1996)

(14) 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.
Adam: and **I**'m going to.
Sherm: yeah. (Grimshaw 1987)

Such shared utterances are not amenable to G&C's standard sentential grammar. Unlike what is licensed in the G&C model, the contextual parameters relevant to such cases need to shift mid-sentence, before whole propositional contents have been derived. Moreover, such role-switches include cases where the same structure can be used both as expressing a speaker's own voice and as a subsequent quotation, as the following show:

(15) 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 [BBC Transcripts, *Jamaica Inn*, Episode 1]

(16) Lucy: Me sister's alive! That girl they found in the field, it wasn't Catherine. She's written to me. She's coming to visit.
Patience: Oh, God, now I'M pissing myself! What? You don't think it was written by a ghost, do you? Or someone pretending to be your sister?
Miriam: That is the nastiest, dirtiest thing anyone has ever done
Patience: says Black Peter's strumpet! What are you crying for? [BBC Transcripts, *Jamaica Inn*, Episode 1]

(17) A: SOMEONE is keen
B: says the man who slept here all night [BBC Transcripts, *The A-Word*]

In all such cases, issues of "footing" (Goffman 1979), namely changes in perspectives and roles assumed by interlocutors, intersect with syntactic/semantic issues of direct/indirect speech forms and speech-act responsibility (Gregoromichelaki and Kempson 2016; Kempson et al. 2017; Goodwin 2007) in ways challenging for orthodox grammatical frameworks like those of G&C, Maier (2014a, b), and Potts (2007).

For these reasons, this paper argues that an adequate account of the function of quotational devices cannot be given within standard NL theories involving the encapsulation and division of labour between semantics/pragmatics and syntax. Instead, it adopts a dynamic, incremental perspective that takes a fine-grained

analysis of joint action among conversational participants as the basis of the definition of the grammar as advocated within Dynamic Syntax (DS, Kempson et al. 2001; Cann et al. 2005; Gregoromichelaki 2006, 2013a) augmented with type-theoretic representations (DS-TRR, Purver et al. 2010; Purver et al. 2011; Eshghi et al. 2015; Hough 2015; Gregoromichelaki and Kempson 2015). As argued within the context of DS-TTR, the grammar itself needs to incorporate incremental interpretation and interaction with the context in order to deal, not only with what have traditionally been analysed as indexicals, e.g. elements like *I* and *you*, but also speech-act information, roles assigned to participants etc., and, most relevantly here, subsentential shifts of the context of evaluation.

The structure of this chapter is as follows: First it is argued that the echoing and metarepresentational abilities that underpin quotation are not peculiar to citation and reported discourse but also occur in conversational interaction, in particular, in cases of repair and, in general, during the process of "grounding", i.e., the signalling of comprehension, correction, (dis)agreement, or clarification of a previous utterance (Sect. 2). We will then examine two recent proposals implementing an appropriate contextual integration of syntax and semantics that will provide ingredients for the present analysis of quotation: (a) The PTT[3] model (Poesio and Traum 1997, 1998; Poesio and Rieser 2010), which implements incremental semantic evaluation by means of subsentential updates of 'information states' (Sect. 3); and (b) Ginzburg and Cooper's model (2014), which reformulates both syntactic analyses and semantics via Type Theory with Records (TTR, Cooper 2005, 2012) (Sect. 4). Exploiting the potential of TTR, G&C attempt to integrate a standard view of quotation within a dialogue model without ad hoc devices that are not independently needed in the analysis of conversation (Sect. 5). It will then be argued that G&C's constructional approach is not sustainable because the lack of incrementality in the grammar deprives this account of the resources needed to deal with data of various intermediate phenomena like free, hybrid, and mixed quotation (Sect. 6). In view of this, an alternative grammar formalism (DS-TRR) is presented that integrates some of the ideas of the PTT and G&C accounts but within a distinct incremental processing architecture (DS) that accounts naturally for the properties of these intermediate phenomena (Sect. 7).

2 Metacommunicative Interaction as the Basis for Quotation

The human ability to employ quotational mechanisms has been proven pernicious for the formal languages of logic which isolate the descriptive function of NLs. For this reason, quotational devices are presented as exceptional by philosophers and logicians, even though they constitute a basic aspect of the ubiquitous potential for reflexivity in human behaviour and thought. Since NLs, and other communicative

[3]PTT is not an acronym but somehow composed out of the names of its inventors.

means, reflect this cognitive potential directly, they cannot be adequately analysed in abstraction from such "devices of quotation". This becomes apparent when one examines the mundane metacommunicative mechanisms required to underpin coordination during conversational interaction. First, there are various frequently used reflexively interpreted elements occurring in all NLs, e.g. indexical pronouns and tenses; such elements require reference to the parameters of the utterance event itself (e.g. interlocutor roles, event time) for their interpretation (see e.g. Recanati 2010). Second, achieving coordination in conversation crucially relies on the ability to express and to perceive certain (aspects of) linguistic signals as pertaining reflexively to the communication process under way. This is not a sophisticated human ability. Even prelinguistic children and non-human animals can engage in activities which involve some kind of play, e.g. pretending to fight instead of fighting (see e.g. Bateson 1987), demonstrating a dance, play-acting etc. Within such framings, the actions that occur carry the implicit meta-message that their usual significance (e.g. hostility, aggression) is suspended and, instead, some other kind of significance needs to be sought. Similarly, during the mundane linguistic activity that occurs in everyday conversation, many speech act(ion)s employ linguistic elements that can function in various ways, instead of being exclusively confined in their supposedly basic referential function. For example, repetition of a phrase can be construed as echoing some just used utterance token, for example in cases of *clarification questions* (see e.g. Ginzburg 2012):

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

(19) A: Did Bo leave?
B: Bo? ('Who are you referring to by your use of the name 'Bo'?'; 'Did you utter the word *Bo*?')

In addition, further phenomena, like correction and related speech amendments (*other-* or *self-repair*), also require similar "echoic" mechanisms. On such occasions, a new token is produced that has to be recognised as similar to an antecedent one. Subsequently, another token might be produced which has to be recognised as offering a replacement for that antecedent token:

(20) A: Bo, (not Bo,)(I mean) Joe, left.

(21) A: Bo, Bo Jones, 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.

Besides these repair mechanisms affecting the standard interpretation of linguistic tokens, conversation naturally offers an environment where the construction, interpretation and authorship of utterances is spread across interlocutors (*split*

utterances, Gregoromichelaki et al. 2011). These are illustrated in turns 3, 4, 5, 12, 14, 21 below (square brackets enclose overlapping speech).

(25)

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 one week and have about a month off or go on the dole for a month.
10. B: So then what happened was, did the Dock Commission say you can't have your own men anymore?
11. A: That's right they had to go on a rota.
12. B: Run by the Dock Commission?
13. A: Run by the Dock Commission. See the dockers then all got together and they said right so many men for that job, so many for that job and that didn't matter who they were, they had to <*unclear*> their job, all the way round the dock.
14. B: Whether they wanted to go on that job or not?
15. A: Whether they want to go or not, they take their turn and the employer had to pay a percentage into the pool what those men earned, so when those men hadn't work at all they drew their money from the National Dock Labour Board.
16. B: Is this where the National Dock Labour Board came into existence?
17. A: That's how how they come into existence, yes <*name hidden*> he was a man what introduced that.
18. B: When was this?
19. A: Oh that's er, I would say about nineteen forty roughly [CLEARS THROAT] I'd 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 [British National Corpus H5H: 89–113]

In many cases of split utterances, as in quotation, the current speaker can be seen as the *animator* (utterer) but not necessarily the *author* or *principal*, i.e., the person bearing the responsibility for the speech act(s) performed (Goffman 1979; McCawley 1999; Antaki et al. 1996). For example, in (25)-4 and (25)-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 A was going to say.

Reported speech is one of the environments where the phenomenon of split-utterances is frequently observed due to the assumed projectibility (high predictability) of the upcoming continuation (Lerner 1991). Additionally, I would argue, such environments also offer the potential for construction of a jointly derived evaluation of somebody's behaviour/thought, a jointly derived plan of action, or, in general, an affiliative opportunity—factors which, in my view, override the normativity of the usual turn-taking strictures (see also (11)–(14) earlier):

(26) 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? (Lerner 1991)

(27) Roger: they rationalized it. they say heh heh heh
Louise: it wasn't there it was all in his imagination. (Lerner 1991)

(28) 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] (Antaki et al. 1996)

(29) 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. (Lerner 1991)

As can be seen from these examples, it is not the case that even clear cases of "reported speech" ((28)–(29)) involve the reproduction ("echoing") of actual utterances/thoughts rather than imaginary or future ones. Instead they function as coordinating devices, e.g. in planning future action.[4]

These types of joint action pattern syntactically either with direct or indirect reports depending on the attribution of authorship to the utterance and correlating with the function it performs. It has been established that indexicals in split utterances are assigned reference according to parameters of context (e.g. interlocutor roles) that shift incrementally during the unfolding of the action and utterance (Gregoromichelaki et al. 2011); see e.g. (30) which is a simple continuation of A's utterance with no verbal echoing intent, even though confirmation might be simultaneously requested as to whether this is an appropriate representation of the semantic content of A's query:

(30) A: Oh, I am so sorry, did you burn
B: myself? No, it's OK / #yourself? No, it's OK

[4]For quotation of thoughts, see Maier (2017).

On the other hand, indexical shifts can reflect subtle distinctions in the function of the current utterance, for example confirmation of various suggestions about aspects of the speech act (in (30), which parallels indirect reports), or clarification of reference (paralleling indirect reports in (31) and direct reports in (32)):

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

(32) A: Did you leave?
B: You? ['Who do you mean "you"?']

Given that all these phenomena—continuations, clarification, and, in general, repair exchanges—are universal and present from very early on in language acquisition, the means and skills involved in the production and comprehension of reported discourse do not appear idiosyncratic or ad hoc. Taking this assumption seriously, two recent holistic models of NL use, HPSG-TTR (Ginzburg and Cooper 2014) and Dynamic Syntax-TTR, seek to model reported discourse via the same mechanisms as those used to analyse such everyday conversational phenomena as those just cited. We turn to these two models next.

3 The Formalisation of Metapragmatic Awareness: Information States and Utterance Events

Recent efforts in formal semantics, inspired by work in Situation Semantics and Discourse Representation Theory (DRT), have shifted attention away from a strict formulation of a truth theory for *sentences* in order to develop theories of semantic interpretation for *utterances* in context. For this purpose, representational systems allowing the specification and seamless integration of multiple types of information have 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 dynamic semantics to take into account 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 psychology (Clark 1996)—that semantic interpretation of utterances relies on the participant's *information state*, i.e., an evolving representation of the context for each participant in a conversation. Such representations also include the reification and explicit representation of the *utterance event/situation*, i.e., the situation providing for the instantiation of the contextual parameters of the conversation itself in order to account for the reflexive reference of indexical elements (see also Maier 2017, this volume).

Even more innovatively compared to previous versions of DRT, in this account, the occurrence of utterances of subsentential constituents is recorded (as *micro-conversational events*) in a certain temporal order so that the gradual accumulation of utterance micro-events becomes part of the structure 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, thus implementing incrementality (see also Larsson and Traum 2000; Stone 2004). As regards the characterisation of particular dialogue actions, in this model speech acts are conceptualised as events too, termed as *conversational events*. This is based on the fact that interlocutors can make metapragmatic statements employing such events as the antecedents of anaphoric expressions:

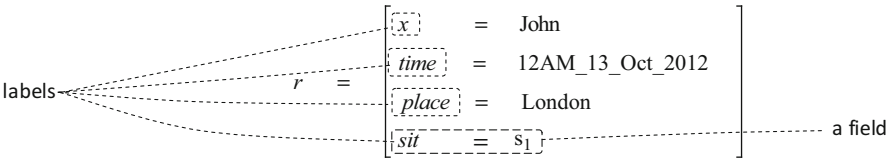
(33) A: You're an idiot.
B: **That** was uncalled for. [that = A's insulting B]

4 Utterance Events in Type Theory with Records

Another recent articulation of the effort to develop theories of semantic interpretation for *utterances* in context (albeit excluding the incrementality dimension),[5] but with richer semantic structures, has been via the development of Type Theory with Records (TTR). TTR provides a general semantic representation format that can integrate both low-level (sub-symbolic) perceptual information (see e.g. Larsson 2011) and high-level conceptual inference enriching the 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 NL use in interaction and the demands this places on appropriate semantic models (see e.g. Ginzburg and Cooper 2004; Ginzburg 2012).

TTR as a representation format provides recursive data structures reminiscent both of the feature structures in Head-Driven Phrase Structure Grammar (HPSG, Sag et al. 2003) and, semantically, of discourse representation structures (DRSs). Records, like *r* below, are structured collections of 'fields' consisting of assignments of entities as values to 'labels', the equivalent of discourse referents/variables in DRT. Each line inside of *r* below is a field, and *x, time, place*, and *sit* are labels.

(34) A record representation:



Such records can then be taken as the representation of events/situations in the world. Importantly, contexts and speech events can be represented as such records

[5] At least not until recently; for some preliminary moves in this direction see Ginzburg et al. (2014).

in order to provide for the instantiation of various contextual parameters as imposed by linguistic forms.

Records (and therefore effectively events/situations) are classified by types which are called *record types*. Unlike the basic Montagovian types, record types (like records) are structured and recursive (i.e. [record] types can be embedded as the value of a label within a record type). Additionally, dependencies can occur among the values assigned to the labels: in many cases the vertical ordering of the fields reflects such dependencies in that if we first introduce a typed entity it can then yield a context for the subsequent introduction of a new type dependent on it. 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 (34) is of the type T in (35) 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 contains evidence that John runs—perhaps s_1 is some actual event of John's running[6]:

(35) A record type representation: T $=$ $\begin{bmatrix} x & : & \text{IND} \\ time & : & \text{TIME} \\ place & : & \text{PLACE} \\ sit & : & \text{RUN (JOHN)} \end{bmatrix}$

Types, which are abstract modelling constructs, can thus be conceived as categorisations of entities and events/situations that provide the interface between the external world and cognition; for example record types, namely categorisations of situations, can be used as structured 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 or 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 (Cooper 2005) and reporting verbs. Moreover, types are modifiable by manipulating their structure, e.g. adding or deleting fields/values. For this reason, the underspecification, enrichment, and general innovation that permeates type judgements are naturally handled, as during language acquisition, knowledge adjustment, conversational coordination and, more pertinently, as we will see now, quoting (some aspects of) another's speech or predicating properties of (aspects of) a grammatical sign.

[6]This characterisation is related to Martin-Loef's "propositions as types" implementation. Hence *run(john)* is a type of events and it correctly classifies events that constitute "proofs" of John's running (see Ranta 1994: chapter 2).

5 Ginzburg and Cooper (2014): TTR-Modelling of Quotation Constructions

The account presented in Ginzburg and Cooper (2014) provides syntactic analyses, denotations and pragmatic constraints for reporting constructions that utilise independently needed syntactic/semantic entities. G&C aim to demonstrate that a dialogical perspective on NL structure and use directly provides the tools to deal with reported discourse via structures and denotations that are already independently motivated for the modelling of dialogue phenomena.

5.1 Grounding and Clarification

Following the model most comprehensively detailed in Ginzburg (2012), the analysis of dialogue involves defining the appropriate updates of richly structured representations of context (*information states*) formalised through TTR. Adoption of the TTR formalism as a uniform representational format allows Ginzburg to model the interactions of the distinct components of the model: the semantic ontology, the grammar, and a system of context updates underpinning the modelling of the interlocutors' common ground.

To be able to include in the model of contextual updates the metacommunicative function of certain utterances, for example clarification requests (5) and (18), dialogue processing is assumed to rely on a process of *grounding* (Clark 1996). Grounding is a necessary dialogue phase during which each participant either confirms that they have understood and agreed with the utterance addressed to them, thus incorporating it in their information state (i.e. their own version of the "common ground"), or they seek clarification of aspects that have not been grounded. Ginzburg formally elaborates the grounding requirement along two dimensions. First, grounding is not monolithic and 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. Second, it is not only semantic content that is recorded and manipulated in the participants' information state (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 (19)–(24) earlier).

Ellipsis resolution for such fragments requires that the grammar be able to express reference to utterance tokens, conceptualised here as "utterance events", that can be grammatically characterised along multiple dimensions. Representation of utterance events as *records* (as in (34)) is employed to serve this role. (Partial) grounding is then formalised through the pairing of an utterance event (a record, a

token) with an utterance type, i.e., a (partially defined) grammatical type (a *sign*[7]) that classifies it. Such signs, i.e. grammatical types recording multiple grammatical properties as defined in HPSG, are expressed in the grammar as record types (as in (35)).

So here a major advantage of the use of TTR becomes evident: the grammar and the conversational update mechanisms are provided with access to both types (record types) and tokens (records) of utterances at the object level so they can both be employed for defining the syntactic and semantic updates that need to be effected. This forms the basis for modelling the metacommunicative or metalinguistic functions of NL elements. For example, it is argued that the clarification request in (36) below, which on the surface simply echoes A's use of *Bo*, can be enriched in various (rule-governed) ways, i.e. it requires disambiguation that can be defined through the formalism. So, to enable the eventual grounding of A's query, the grammar needs to specify a reading for the fragment *Bo?* which queries which individual named *Bo* A was referring to in the previous utterance ('intended content reading'). This needs to be distinguished from a separate reading regarding what *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 (36). These readings also need to be distinguished from other readings such as the confirmation readings in (c):

(36) 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 (36)b(i) 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*. Additionally, reference to grammatical types allows for the metalinguistic flavour of clarification queries regarding the meaning of particular utterance tokens as in (36)b(ii).

In order to formally license such constructions and model their disambiguation, HPSG-TTR incorporates a constructional version of HPSG, also expressed in the uniform representational framework of TTR. The rich type theory thus included then allows for the definition of entities modelling both utterance tokens (events as records) and their characterisation via utterance types (signs, grammatical types)

[7]Note that this use of the term 'sign' does not coincide with that of Cappelen and Lepore (2007: chapter 12).

that the grammar and the model of the context can manipulate and reason over. We are going to see how these are employed in the analysis of quotation and reported speech next.

5.2 Locutionary Propositions and Abstract Semantic Objects

Two components in the modelling of the process of grounding are relevant for the analysis of quotation and direct/indirect discourse that concerns us here: (a) locutionary propositions, and (b) abstract quasi-propositional semantic objects assigned as contents to sentential units in order to serve as the arguments of speech act predicates.

5.2.1 Locutionary Propositions

For an utterance in dialogue to be grounded, first it has to be parsed and understood correctly. In HPSG-TTR, the successful 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 TTR-types, i.e., representations that include labels characterising phonology, syntax, semantics and contextual 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.

5.2.2 Abstract Quasi-propositional Objects

In HPSG-TTR, the grammar is assumed to assign to every utterance of a root sentence a speech-act characterisation. During the grounding process, this enables reference to be made to the particular speech act performed by the previous interlocutor, e.g. modelling interpretations like 'Are you asking q' and 'Are you asserting p?'. The (quasi-)propositional arguments p or q in such speech-act specifications are 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 is compatible with each such specification and which is attributed to the agent of the speech act. So, a *proposition* will be what an

agent *Asserts*, 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 defined abstract entities, locutionary propositions (i.e. pairings of utterance events with utterance types), and abstract (quasi-)propositional semantic objects, having been independently postulated for the explanation of NL use in conversation, especially metacommunicative interactions, are taken by G&C to naturally extend to pure, direct, and indirect quotation as we will see now.

5.3 Pure Quotation[8]

Having assumed a constructional version of HPSG, in extending these conversational mechanisms to reporting discourse, G&C define *constructions* for various quotation phenomena that specify the interacting syntactic, semantic and pragmatic constraints for their licensing. According to various theories of quotation (identity theory Washington 1992, proper name theory Tarski 1993; Quine 1940, description theory Geach 1957), the quotation marks somehow modify the reference of the expression that is enclosed within them, so that the expression's usual reference is suppressed and the expression itself somehow is referred to. However, as G&C argue, the analysis of pure quotation requires formal explication of the notion 'expression' (see also Saka 1998; Cappelen and Lepore 2007). The HPSG-TTR model offers a specification of this notion via the invocation of a particular grammar that supplies the TTR-expressed grammatical type characterising such token expressions (see also Maier 2014a,b; Potts 2007).

So, in analysing pure quotation, G&C introduce the assumption that, in general, the contextual parameters included in the information state have to include a parameter Γ that refers to the particular grammar licensing the type of the expression used. Γ roughly corresponds to a formalisation of what in Recanati (2010) is characterised as a shiftable "language". It is a welcome innovation of the view that G&C implement that NL *grammars* are considered as collections of resources for building context-appropriate ad hoc *languages* (syntax-semantics specifications) (Cooper 2012; Cooper and Ranta 2008). Γ therefore will be instantiated in each particular context as the particular subcollection of grammatical resources relevant to the licensing of an utterance event under a grammatical type (sign) in that context (such an instantiation is argued to be a "pre-semantic" process in Recanati 2010; Maier 2014b).

Under this view of the flexibility of grammatical resources, pure quotation can be characterised as invoking a particular set of such resources Γ which provide for its properties as a sign. Pure quotation constructions are usually assumed to

[8]The term *pure quotation* here follows G&C's usage, namely it refers to *citation*. For objections to this usage, see Saka (2013: 939).

occupy an NP/DP position and introduce a referential (singular) term that refers to the linguistic material enclosed in the quotation marks. Accordingly, G&C formally define a special grammatical construction for pure quotation. This construction licenses an NL string, e.g. *You love John*, to syntactically project a phrase (NP) whose semantic content is its grammatical type (sign) relative to a particular grammar Γ. Aspects of this grammatical type, e.g. the phonology, can be picked up in a case like:

(37) 'You love John' starts with a consonant.

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 and the fact that the semantic content of the "construction" is now its grammatical type, instead of its usual compositionally derived reference, explains the opacity of such uses (e.g. *you* in (37) does not refer to the current hearer). According to G&C, use of quotation marks in written discourse indicates this shift of content for such uses of NL elements.

5.4 Direct and Indirect Quotation

Turning to direct and indirect discourse, G&C analyse these as involving two components: (a) lexical entries for the reporting verbs (*quotative predicates*) and (b) constructions that specify the presumed idiosyncratic properties of each such structure. Quotative predicates can select for syntactic constituents denoting (a) locutionary propositions (see Sect. 5.2.1) in direct reports, or (b) quasi-propositional abstract entities (see Sect. 5.2.2) for indirect reports. Both direct and indirect reports are analysed as constructions that involve the combination of a reporting verb, like *say* or *ask*, with a clause whose denotation involves such an abstract semantic object.

A *direct-quotation construction* involves, first, the projection of a *direct-quotational phrase* from the quoted material. This phrase can then serve as the complement of a reporting verb specified to require the type of such a complement. The derived semantic content of a direct-quotational 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. 5.2.1). The utterance event component of such a proposition represents the reported utterance event. This reported event is now associated with the 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 from that of the original reported event or other modifications the reporter might effect (and still be counted as direct quotation).[9] Since the grammar manipulates TTR types as well as tokens, it now becomes possible to express how the original reported utterance event and the reporting event are deemed to be

[9]Recanati speaks of the "language" of the context (2010: 190).

"similar" in some relevant respects (for these, see e.g. Clark and Gerrig 1990). The grammatical type assigned to the utterance event included in the locutionary proposition 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 exceeded threshold value on a similarity measure between the grammatical types of the original and the reporting events. Further, via this construction, the contextual parameters of the standard use of the sign are discarded, as we also saw in pure quotation cases earlier. However, for direct quotation, at the phrasal level, a new set of contextual parameters is introduced now relative to the reported event. In this way, the full content of the reported sign becomes available, unlike cases of pure quotation (where there is no reported event). This allows for the explanation of cases of anaphoric reference to the *content* of the quotation subsequently, as in the Partee examples (such as (4) earlier) where anaphora/ellipsis resolution relies on semantic/contextual processing of the quoted material.

A further innovative advantage offered by this analysis is that by analysing direct quotation complements as denoting locutionary propositions, which include as one of their components a sign (a grammatical type), we can explain the fact that a single sentence can contain predications that address both type and token aspects of the quotation, e.g. use the same quotation as both the complement of a direct-quotation construction and as a pure quotation, as in (38)–(39):

(38) "Was I snoring" *was asked by Bill* and *is a frequently used interrogative clause*.

(39) Bill asked, "Am I snoring?", a sentence frequently used by men who don't think they snore. It is usually answered by "You were before you woke up". (adapted from G&C)

The direct-quotational phrases whose properties we have just described appear as the complements of verbs that combine with direct quotations, for example independent clauses (non-clausal complements for such verbs are also possible). 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 as combining with complements that can have either of 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 (i.e. a direct-quotational phrase, as discussed above). As we've just seen, due to the direct-quotational phrase specifications, the utterance event will be the 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. 5.2), the speech-act characterisation of the original reported event can be available through the abstract semantic object associated with it. Due to this fact, the content of the main clause is now identified with the speech-act content introduced in 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*.[10]

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 of 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, in indirect report cases, the reported event parameters cannot affect the current contextual parameters. The only restriction is that the complement has as its content an abstract semantic object *q* of type *question* (e.g. that derived from *whether Bill 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 through the use of *ask*.[11]

This account is designed to capture the commonalities of direct and 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,[12] even though the structure with the direct-quotational 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 (40) and (41) below entail (42), which is explained because, as we just saw, the contents eventually assigned to the clauses built on the basis of the two versions of *ask* are identical:

(40) Zohar asked whether she snored.
(41) Zohar asked 'naxarti?'.
(42) Zohar asked a *question*, a question about herself.

So the G&C account successfully captures various properties of reporting constructions via the attempted identification 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 (38)–(39) where a single predication can simultaneously address 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 (40)–(42) earlier.

From the present point of view, there are some problems with this latter claim, stemming from the fact that the grammar necessarily associates conventionalised speech-act specifications with each main clause. For the same reason, in combina-

[10]Note that to this a new speech-act specification will eventually be added to the effect that the final (schematic) content will come out as *Assert(Speaker, (Ask(x,q))* for a (schematic) sentence like 'John$_x$ asked q'.

[11]The eventual content derived will again be *Assert(Speaker, (Ask(x,q))*.

[12]As a reviewer notes, there are various other subcategorisation possibilities for such verbs which are not discussed here.

tion 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. The main technical and conceptual reasons for this are, first, 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 shortcomings, in the next sections we will examine these problems to motivate the claim that we need to introduce an alternative account that builds on some of G&C's insights but is formulated within an incremental, dynamic framework, namely DS-TTR.

6 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 is syntactically employed as an NP (or DP in some syntactic frameworks) and semantically plays the role of a singular term. The G&C account is explicitly addressed to such closed quotation cases only. However, I believe that G&C have provided some of the resources that make a more inclusive account available, i.e., covering also the phenomenon of *open quotation*, where the quoted material is not integrated in a clause. In my view, the only factor that prevents the integration of such phenomena in the G&C account is the standard assumption of an independent syntactic level of analysis in the grammar and the lack of 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 2014a) from dealing with the whole range of data as we will now see.

The G&C constructional account inevitably adheres to the standard strict division between direct and indirect quotation. However, this strict distinction can be disputed as there is a host of phenomena that lie in a continuum between these two supposed extremes. First, there are languages where there is no such strict distinction; instead various syntactic features of the utterance can indicate either more or less syntactic/semantic integration, for example combinations of complementiser, mood change, or verb-second:

(43) (a) Peter sagte, dass er das nicht machen könne. [subjunctive + complementiser]
Peter said that he that not make can.SUBJUNCTIVE
Peter said that he couldn't do this
(b) Peter sagte, dass er das nicht machen kann. [indicative + complementiser]
(c) Peter sagte, er könne das nicht machen. [subjunctive + verb-second]
(d) Peter sagte, er kann das nicht machen. [indicative + verb-second]
(German, from Brendel et al. 2011).

In my view, such phenomena show that the grammar needs to provide mechanisms for processing rather than rigid constructional analyses.

Next, there is always the possibility of introducing quotational elements, for example elements extraneous to the reporter's dialect, within a report otherwise characterisable as indirect (and without the use of [potentially invisible] quotation marks, contra Maier 2014a):

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

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

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

Free indirect discourse Banfield (1973) is similar to indirect reporting in that there is shift of tenses and indexicals. However, usually, there are no overt reporting indications, temporal adverbials are evaluated with respect to reported event, 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 reporting context (see e.g. Eckardt 2014):

(46) **Mary** felt relieved. If Peter came **tomorrow**, **she** would be saved. (Recanati 2000)

(47) **Marie** was wondering. Did **her** brother arrive? (Bonami and Godard 2008)

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

(48) The baritone was asked **what did** he think of Mrs Kearney's conduct. (Koder and Maier 2014: fn. 1 citing McCloskey 2006, from James Joyce, *Dubliners*)

These phenomena cannot be handled by the G&C account because their model requires the grammar to deal with phrasal constructions that specify either direct or indirect features. In all these cases, however, as in pure quotation, there is no necessity for a reporting verb to determine the appearance of a quotational interpretation. Another phenomenon that is excluded for the same reasons is that of *mixed quotation*, a combination of direct and indirect discourse, characterised 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; De Brabanter 2010a):

(49) 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 is presented as having the same content as what the reported speaker said. But, as in direct reporting, there seems to be indication that the

reporting speaker, or some other speaker (Recanati 2010), used similar words as those appearing in quotation marks. Since both these aspects of such 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, because such structures in G&C's account would have to be analysed through the constructions defined for the reporting verbs, which unify the subject of the report with the utterer of the quotation, the option of such quotations echoing other speakers/thoughts is excluded.

Along with the G&C account, all standard syntactic/semantic and pragmatic models face problems accounting for certain alleged peculiarities of mixed quotation. First of all, like direct quotation and, as we saw earlier in (44), even with indirect quotation, there is the possibility to shift not only the interpretation of indexicals but even language in the midst of reporting such utterances:

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

(51) A doctor tells him [Gustave Flaubert] he is like a 'vieille femme hystérique'; he agrees. (De Brabanter 2010a, from *TLS Online*, 18 December 1998)

Another issue that arises for formalisms that do not embrace the incrementality of processing in the grammar but, instead, attempt 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:

(52) She replied, 'I live alone. My son lives alone too. We both prefer it that way'. (De Brabanter 2010a, from Huddleston and Pullum 2002: 1026)

(53) 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". (De Brabanter 2010a, from *Independent Arts*, 17 September 2004)

(54) David said that he had donated "largish sums, to several benign institutions". (Abbott 2005)

(55) Mary allowed as how her dog ate "odd things, when left to his own devices". (Abbott 2005)

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

(57) [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 [. . .]. (De Brabanter 2010a, from *The Guardian*, 20 March 2005)

This cannot be handled by a grammar that requires phrases to be built out of conventional constituents that just shift interpretation as they project according to distinct pre-defined constructions. As De Brabanter (2010a) 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 points out, such phenomena have truth-conditional

effects, as can be seen from the distinct interpretations obtained when the quotation marks are removed[13]:

(58) Paul says he's due to present his work in the "paper session". [Paul calls "paper session" the 'poster session']
(59) Paul says he's due to present his work in the paper session. (Recanati 2010)
(60) James says that "Quine" wants to speak to us. [James thinks that McPherson is Quine]
(61) James says that Quine wants to speak to us. (Recanati 2010)
(62) Nicola believes that her father is a "philtosopher".
(63) Nicola believes that her father is a philtosopher. (Cappelen and Lepore 1997)

This clearly indicates that a grammar formalism needs to integrate interaction with context 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, similarly to the G&C account, by assuming that the contextual parameters that need to be included in a grammatical analysis must represent various sources for the grammatical resources employed, for example entities like potential idiolects, dialects, and languages. It seems to be the case that we need a rather liberal characterisation of such entities since the processing devices involved (the 'grammar' from our point of view) in such uses are open-ended and are not dependent on any actual folk-linguistic characterisation as the examples in (58)–(63) show. Cases involving use of quotation marks as indicating the speaker's dissociation from some usage of words ('scare quoting') can also be accounted for through such grammar-shifts (see Sect. 7.3.3). The pragmatic process leading to the (local) shift and the instantiation of such variables can be conceptualised as described in Saka (2003/2005: Sect. 3.1).[14] Other such phenomena can be treated in addition as potentially 'echoic' in the sense that the contextual parameters will also include a reported utterance- or thought-event (whether actual or generic, habitual etc.; see also Predelli 2003, Maier 2017). In my view, all such analyses are available and compatible (contra Saka 2003/2005) if the grammar provides fine-grained mechanisms rather than static characterisations of "expression types". However, for such analyses to be available, it is crucial that such shifts of contextual parameters be available subsententially during the interpretation of a fragment of the utterance being processed. Such parameters, additionally, should neither necessarily project syntactically nor be defined only at the root level (as in many standard grammatical/semantic frameworks).

In this connection, another related issue that arises for the G&C account is the fact that the speech-act specification associated with each main clause is taken

[13]This does not imply that quotation marks are necessary for such uses. It is just that they constitute signals that facilitate the alternative interpretations indicated (see also Saka 2003/2005). In a particular context of use, their presence is not required for such interpretations to be possible.

[14]Such a process does not have to be conceived in a Gricean manner as inference driven by the need to derive the speaker's intention. Instead it can be implemented mechanistically in the grammar in the sense, explained later, that the hearer understands the speaker's actions through mirroring these actions as specified by her/his own grammar (see Gregoromichelaki et al. 2011).

to be conventionalised, i.e. there is a selection from among a predefined set of such illocutionary forces (see earlier Sect. 5.2.2). However, what precise speech-act specification is potentially assigned to each utterance is an open-ended issue and subject to contextual determination so that there can't be any default specifications determined by the grammar (Gregoromichelaki and Kempson 2015); the grammar just needs to include mechanisms for such optional pragmatic determination to potentially affect truth-conditional content on the way to deriving contents for the full utterance. Support for this claim is provided by the fact that indirect report complements can appear with a multitude of speech-act denoting framing verbs (and this class of verbs is open-ended):

(64) 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." [*Cyprus Mail*, 31/05/11]

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

(65) 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."
(66) Mario Draghi clarified: "There is no practice or methodology for a country to exit the eurozone."
(67) Mario Draghi offered a clarification of his previous statements.

Such alleged "entailments" are not qualitatively different from the ones offered by G&C in (40)–(42). However, they cannot be explained as arising from a range of fixed speech-act specifications and special semantic objects defined in the grammar, which is what provides the explanation of (40)–(42) in the G&C model. If there is a mechanism for deriving the inferential pattern in (64)–(67) pragmatically, it can also be used to derive the inferences in (40)–(42) as long as such pragmatically inferred contents can interact with grammatical specifications at an appropriate level.

On the other hand, the alleged inviolable restrictions implemented for indirect reporting in the G&C account and others do not hold for *mixed quotation*, a construction structurally similar to indirect quotation. 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 reporting verb (Geurts and Maier 2005; Cumming 2005; Anand and Nevins 2003) as in direct quotation structures:

(68) Bill Watterson said that reality "continues to ruin my life". (Maier 2014a)

However, *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), so that any actual such constraints have to be implemented elsewhere:

(69) Quine remarks that quotations have a "certain anomalous feature" that "calls for special caution"; (Davidson 1984: 9)

(70) Who did Mary say that she would "never misunderestimate ever again"? (Maier 2014a)

Regarding the lack of syntactic opacity in mixed quotation, Maier (2014a) claims that certain features of the quoted original in mixed quotation obligatorily have to be adjusted to fit the new quoting environment. For example, he claims (citing Shan 2011) that the grammatical gender agreement displayed by a quoted phrase in gender-determining languages has to be adjusted to fit its new environment. However, this is not an absolute constraint either but a choice concerning whether the quotation echoes faithfully the form of an utterance or not. For example, there are cases like (72) where this alleged restriction does not hold because the incompatibly gendered characterisation (as shown in (71)) happens to convey exactly Maria's words:

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

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

In conclusion, these intermediate phenomena—free, hybrid, and mixed quotation—show that there is no strict distinction between direct and indirect reporting so that there is no need for distinct phrasal 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. Instead, as in mundane conversational interaction, it also has to be assumed for quotation that the grammar provides fine-grained mechanisms according to which speakers/writers can freely shift the mode of presentation and perspective of their utterance, indicate who takes responsibility for its content and form, or draw attention to some of its properties at any sub- or supra-sentential level. This argues against a model of NL-grammar that ignores the psycholinguistically established incrementality of processing and the dynamic nature of context updates. On the other hand, it provides support to the claim that grammatical semantic/syntactic constraints are not qualitatively different from pragmatic processing, and, therefore, cannot be segregated in a distinct abstract static model that provides analyses only for linguistic strings. This is shown most clearly by the fact that contents provided by NL-utterances can compose with a variety of demonstrating events, like gestures, noises, or pictorial signs in written language:

(73) The car engine went [BRMBRM], and we were off. (Clark and Gerrig 1990)

(74) The boy who had scratched her Rolls Royce went [RUDE GESTURE WITH HAND] and ran away. (Recanati 2010)

(75) Every person who went [DEMONSTRATION OF RUDE GESTURE/BRMBRM] was arrested. (adapted from Postal 2004)

To capture such phenomena and desiderata as an intrinsic consequence of the framework, we now turn to a grammar formalism that takes into account the fact that NL is primarily a form of action, produced and interpreted in context in a time-linear manner (for similar intuitions in the quotation literature see Saka 1998, 2003/2005). The next sections aim to show that the data mentioned above, which are highly problematic for other formalisms, find natural explanations from such a perspective.

7 Dynamic Syntax

In distinguishing between open and closed quotation (see earlier Sect. 6), Recanati (2010) makes an alleged important distinction: open quotations are primarily "demonstrations", 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, emphasis mine)

Closed quotations in contrast, according to Recanati, carry additional referential meaning due to their integration in the linguistic system. From that point of view, this distinction reflects the standard conception of NL-analysis as requiring a specifically linguistic grammar on the one hand and a separate component of pragmatic inference, concerned with human action, on the other. In contrast, the framework adopted here, Dynamic Syntax (DS, Kempson et al. 2001; Cann et al. 2005; Gregoromichelaki and Kempson 2013, 2015, 2017), presents a more radical alternative concerning the status of the syntax/semantics components of the grammar and their integration with pragmatics. Under this conception, syntax is not a level of representation at all but a set of more-or-less domain-general routinised mechanisms (packages of *actions*) for integrating or producing communicable signals, with the grammar standing in continuity with other categorisation processes of intentional/non-intentional stimuli. From this point of view, for DS, 'demonstration' (whether echoic or not), not 'reference', is all there is in linguistic processing in general so the opposition 'closed' vs 'open' quotation cannot be adopted (for a similar but less radical conception regarding syntax see also Saka 1998, 2003, 2003/2005). Nevertheless, Recanati's insight that 'closed quotation', like other non-linguistic demonstrations (see (73)–(75)), can be recruited as linguistic constituents can receive natural expression in DS as we are going to see in Sect. 7.3.1.

DS, instead of conceiving of NL as a code licensing form-meaning correspondences, models the mechanisms of processing, conceived as (epistemic) act(ion)s interlocutors engage in during the production and comprehension of both meaning and forms. So all levels of traditional NL analysis are reconceptualised as actions performed and assigned meaning within a context. In this respect, DS can be seen as a psycholinguistically inspired 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 constituency as a level of explanation. Instead it implements the assumption that grammatical constraints are all defined procedurally. Such constraints guide the progressive development of conceptual representations along with contextual information ('information states'), with partial interpretations and strings emerging step-by-step during social interaction on a more or less word-by-word basis. In the model adopted here (DS-TTR, Purver et al. 2010), Dynamic Syntax is enriched with conceptual representations formulated in the Type Theory with Records framework (TTR, Cooper 2005, 2012; see earlier Sect. 4). TTR is able to integrate information from perceptual and subsymbolic sources (Larsson 2011, 2015), which captures directly the fine-grained dynamics of dialogue, its potential for integrating input from various modalities under a single processing mode, and the potential for underspecification and enrichment (Purver et al. 2011; Eshghi et al. 2015). Thus DS-TTR is formulated as a system which crucially involves:

- An action-based architecture (DS) that dynamically models the development of unitary TTR representations (information states) integrating multiple sources of contextual information
- Word-by-word incrementality and predictivity within the grammar formalism
- Parser/generator (i.e. speaker/hearer) mirroring and complementarity of processing actions as part of the grammar.

This perspective, when applied to dialogue modelling and quotation devices, sheds new light on the phenomenon of split utterances seen earlier in (25), taken up below in Sect. 7.2; how the mechanisms apply there, in combination with some of the tools provided by the G&C account, allow for modelling the continuity of mechanisms underpinning pure quotation, direct and indirect discourse, and mixed/scare quotation (as we will see in Sect. 7.3). Since both dialogue phenomena and reporting/citation devices are using the same grammatical resources they are predicted to interact. This is shown with quotation data which receive analysis with the same means as dialogue phenomena.

7.1 Incrementality/Predictivity and Radical Context-Dependency in the Grammar

Instead of deriving sentence structures paired with propositional meanings, as in models of *competence*, the DS formalism models directly the interlocutors' *performance* in processing word-by-word NL strings and meanings in interaction with the non-linguistic context. For NL use in conversation this is a crucial explanatory factor since many of its metacommunicative features rely on such incremental production and comprehension. For example, the frequent occurrence of constituent clarifications (see earlier (5), (19)) in conversation shows that utterances can be

processed and understood partially without having to map a sentential structure to a full proposition. Moreover, the process of *grounding*, invoked and modelled at the propositional level by Ginzburg (2012) (see Sect. 5.1 earlier) relies on the appropriate positioning of items like inserts, repairs, and hesitation markers, a positioning which is not arbitrary but systematically interacts with grammatical categories and derivations at a subsentential level (see e.g. Clark and Fox Tree 2002). During grounding, addressees display their comprehension and 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 feedback received from hearers as their turn unfolds (Goodwin 1981). Hence the grammar must be equipped to deal with such metacommunicative signals in a timely and integrated manner, namely by incrementally providing online syntactic licensing, semantic interpretation, and pragmatic integration. In addition, the turn-taking system (see e.g. Sacks et al. 1974) seems to rely on the grammar, as it is based on the predictability of (potential) turn endings in order for the next speaker to time appropriately their (potential) response; in this respect, experimental evidence has shown that this predictability is grounded mostly on syntactic recognition rather than prosodic cues, intonation, etc. (De Ruiter et al. 2006).

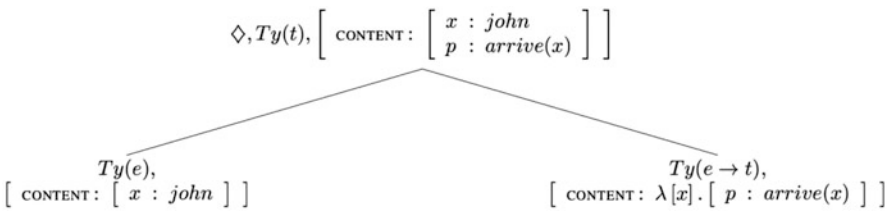
For all these reasons, the DS-TTR model assumes a tight interlinking of NL perception and action by imposing top-down predictive and goal-directed processing at all comprehension and production stages so that input and feedback are constantly anticipated by relying on contextual linguistic and non-linguistic information in order to implement efficient performance. Concomitantly, coordination among interlocutors can then be seen not as inferential activity but as 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 (Gregoromichelaki and Kempson 2013). These actions perform step-by-step a mapping from perceivable stimuli (phonological strings) to conceptual representations or vice-versa. Production uses the testing of parsing states in order to license the generation of strings while comprehension invokes prediction of upcoming input in order to constrain efficiently the usual overwhelming ambiguity of linguistic stimuli.

7.2 Conversational Phenomena in DS-TTR

In DS-TTR, the conceptual contents derived by processing linguistic strings are represented as trees inhabited by record types (see earlier Sect. 4 and (76) below). The nodes of these trees are annotated with terms in a typed lambda calculus, with mother-daughter node relations corresponding to predicate-argument structure (by convention arguments appear on the left whereas predicates appear as the right daughters). Abstracting away from details for now, for example the content associated eventually with the string *John arrives* will be the functional application of the lambda term $\lambda x.Arrive'x$ inhabiting the function daughter, to the singular term

derived by processing the name *John* (in TTR terms, notated as the sole witness x of type *john*, x *: john*, in the display below). $\lambda x.Arrive'x$ has the semantic type of a one-place predicate, which in the logic and diagrams is shown as: *Ty(*$e \rightarrow t$*)*). The result of functional application will be a propositional type (*Ty(t)*) (the witness exemplifying such a propositional type will be an event/situation, notated as p in the display below). For simplicity we assume that *John* will trigger the search of the context for an individual (of semantic type *e; Ty(e)*) named 'John'.[15] In terms of representations, such contents are accumulated in *fields*, recursive label-value pairs, of TTR record types (see earlier (35) in Sect. 4). Labels (like p or x below) stand for the witnesses of the types expressing derived conceptual content. The semantic content is accumulated as the value of a designated content field (indicated as CONTENT in the simplified diagram below):

(76) (simplified) DS-TTR representation of the conceptual structure derived by processing *John arrives*:



A pointer, ◇, which moves around the tree nodes as the result of defined language-specific processing actions (thus accounting for word order), indicates the current node of processing, the current locus of attention.

Words and syntactic rules are conceptualised in DS-TTR as *lexical* and *computational actions* respectively, i.e., as triggers for inducing packages (*macros*) of atomic actions if certain specified conditions are satisfied at the current locus of attention (the **IF** specification in (77) below). Such actions include the triggering of contextual searches for conceptual content (`find`, (`fresh`)`put`, `substitute`), building conceptual tree-structure (`make`), copying values, introducing predictions of upcoming input, or, finally, aborting (`abort`) in case the conditions of use of the word/rule are not satisfied in the current linguistic and non-linguistic context. In this sense, words and rules can be seen as 'affordances', i.e., possibilities for (inter)action that agents attuned to these possibilities can recognise, predict, and manipulate (Gibson 1977). For example, the simplified lexical entry for *arrives* shown below first checks whether the pointer is at a node predicted to be of predicate type (indicated as: *?Ty(*$e \rightarrow t$*)*) and, if this condition is satisfied, it introduces, via the execution of the atomic action `put`, the conceptual content represented by the function $\lambda x.Arrive'x$ (the full specification of the macro includes the execution of

[15]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, namely terms carrying uniqueness implications and descriptive content. Here no stance is taken on this issue as it does not affect current concerns.

further actions relating to tense, mood, agreement etc. through the employment of a set of actions like `make`[node], `go`[to node], `abort`[processing], etc.):

(77) Lexical entry for an intransitive verb:

arrive:

IF	$?Ty(e \rightarrow t)$
THEN	$\texttt{put}(Ty(e \rightarrow t))$
	$\texttt{put}(\lambda \left[\begin{array}{l} x : e \end{array} \right] . \left[\begin{array}{l} x : e \\ p : arrive(x) \end{array} \right])$
...	
ELSE	`abort`

Following Gregoromichelaki (2006), it is also assumed that a propositional representation (of type *t*) always includes an indication that events/situations belong to some world/time of evaluation (Recanati 2004)[16] as a contextually derived value of specified parameters. Additionally, each predicate type derived as the CONTENT field at each subnode of the tree includes independently shiftable world/time/situation parameters to account for well-known cases of differentiation among the parameters of evaluation for various predicates in a sentence:

(78) The fugitives are now in jail. (Enç 1986)

All such context-dependent values are derived through the fact that various linguistic elements are defined as initially introducing *metavariables* in the conceptual representation. *Metavariables* in DS (indicated in capitals and bold font) are temporary place-holders introduced to enforce their later substitution with values (variables or constants) from the current context. For example, pronouns, anaphors, ellipsis sites (auxiliaries in English), tenses, modal verbs, etc., lexically introduce metavariables of various types (of type *e* or predicative types) and restrictions constraining their subsequent replacement by values derived from the linguistic or non-linguistic context.

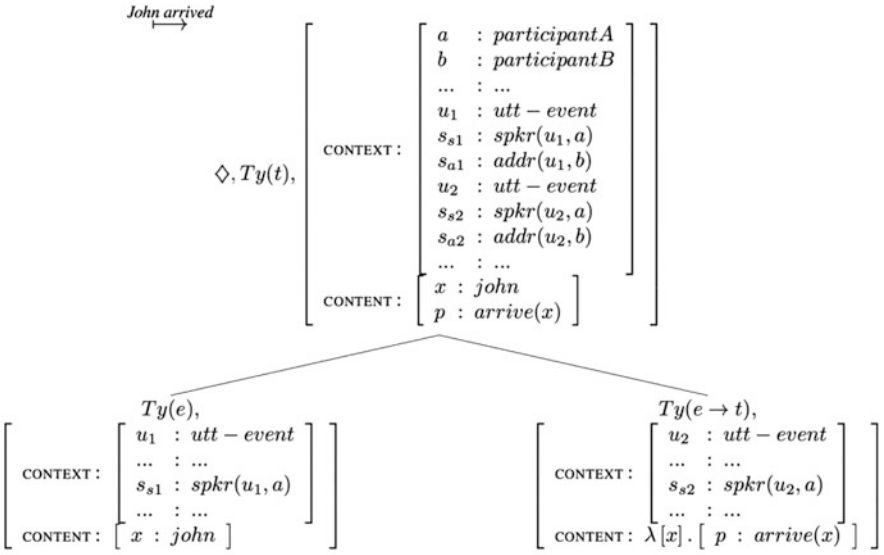
Consequently, in order to model the interpretation process of reflexively interpreted elements like indexicals *I*, *you*, and *now*, contextual parameters regarding the utterance event are recorded in a structured CONTEXT field[17] on which the

[16]For eliminating worlds from the semantics, replacing them with more psycholinguistically plausible TTR contents in terms of (types of) situations, see Cooper (2005). Here we maintain the more conservative view for brevity of exposition.

[17]The differentiation CONTEXT vs. CONTENT fields is for exposition purposes only, just for the convenience of shortening reference to fields in the displays; it does not signify any substantial claim regarding any qualitative differentiation among the parameters handled. In TTR there is always an intuitive inclusion of the context in that, via the notion of *dependent types*, subsequent fields can depend on elements introduced previously (up along the vertical dimension in record types) but not the other way round. In terms of expressivity, reference to a value in some record (type) can be indicated via the definition of *paths* leading to specific values; we show such paths with dots separating the sequence of steps, for example $r.l_1.l_2$, refers to the value of label l_2

CONTENT field depends. The CONTEXT field records the occurrence of each word-utterance event (utterance action, notated as e.g. u_1,u_2,...,u_n), including the 'words'[18] that have been uttered, the agent (i.e., the utterer [*spkr* on the diagrams], which can be distinct from the agent taking responsibility for the illocutionary act), the addressee (*addr*), time/location of the event (following the specification of *micro conversational events* in Poesio and Rieser 2010; see earlier Sect. 3), the world parameter of the context, and various constraints in the relations among these terms[19]:

(79) (simplified) DS-TTR representation for *John arrived* with contextual parameters:



Processing of a contextually dependent element, e.g. an indexical pronoun like *I*, first checks whether the pointer appears on a node predicted to be of type *e*

which provides the value of l_1 in record (type) *r*. In the displays here, the various fields are freely simplified and condensed in various ways for uncluttered illustration of the relevant points.

[18]Note that 'words' in DS-TTR are conceptualised as phonological/graphemic/signed shapes, i.e. stimuli that serve as the triggers for DS-TTR actions; not, as usual, 'signs', or 'expressions' (Cappelen and Lepore 2007), or phonology/syntax/semantics feature bundles (cf. Saka (2011) for discussion about the nature of linguistic elements, leading to distinct conclusions).

[19]The initial arrow carrying a word string illustrates the process of *scanning*, the process of recognising stimuli as triggers of lexical macros. Subevents are sequentially numbered through subscripts and further subscripts can be used for mnemonic purposes (the subscripts *s*, *a* here stand for *speaker*, *addressee* but will not be maintained further to avoid confusion with occurrences of subscripts *s* on types where they indicate the subtype of type *e* (entities) that are situations (type e_s).

(anticipated but not yet realised 'goals' are indicated with a ? in front of the expected specification). Then, if this condition is satisfied, an appropriate parameter in the CONTEXT field will be located (the entity that is the speaker) and its value will be copied as the value on the current node, namely in this case, the current speaker value and the indication that it is of *Ty(e)*[20]:

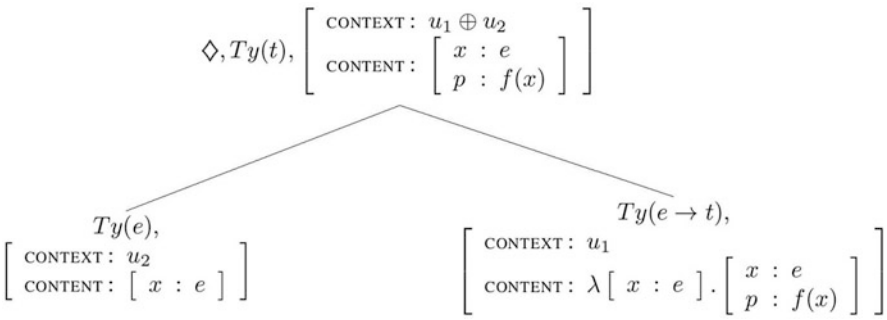
(80) Lexical entry for indexical pronoun *I*:

I:

IF	$?Ty(e), [\ \text{CONTEXT} : [\ s_s : spkr(\mathbf{u}, \mathbf{x})\ ]\ ]$
THEN	$\texttt{put}(Ty(e))$
	$\texttt{put}((\mathbf{x}))$
ELSE	abort

Interspersed with lexical entries, general computational rules can apply without 'scanning'/generation of linguistic input. For example, computational rules induce the concatenation of word-utterance subevents (indicated as $u_1 \oplus u_2 \oplus \ldots u_n$) producing cumulative utterance events at mother nodes of the tree structure. Such concatenation is effected in parallel with the computational actions performing functional application on content-complete nodes (see Purver et al. 2010; Purver et al. 2011 for details):

(81) Concatenation of subevents in the CONTEXT and parallel function application in the CONTENT fields:



For our purposes, we note that there can be additional world and event parameters in the CONTENT field, introduced via the actions of linguistic operators like tense, modality etc., with accessibility relations represented as TTR-dependencies among CONTENT and CONTEXT fields (to deal with such phenomena where shift of

[20]Bold lower case variables in the lexical macros indicate rule-level variables that unify with specified values on the current tree descriptions (parse states). These values are then used in the further execution of the macro (for formal explication see Kempson et al. 2001: 90–91, 311).

evaluation occurs, e.g. conditionals, see Gregoromichelaki 2006; for an alternative TTR formulation that does away with worlds in favour of types of situation, not adopted here for simplicity of exposition, see Cooper 2005). Such potential is needed, independently of quotation, in dialogue since shifts and interactions of contextual and evaluation parameters can freely occur subsententially:

(82) Nun: I'll telephone the Wing Governor. Surely she will appreciate the hiatus in care that has occurred.
Nurse Franklin: Of course she must! In terms of women's healthcare, we're in charge, so **I** wouldn't mince **your** words. [BBC Transcripts, *Call the Midwife*, Series 3, Episode 3][21]

(83) Stanley: Louis, **I** just didn't think
Louis: **you**'d ever hear from me? [BBC Transcripts, *Dancing to the Edge*, Episode 5]

In these CONTEXT parameters, following Ginzburg and Cooper (2014), we now add an NL-use parameter for each sub-event, indicated initially as a metavariable **G** of type *linguistic usage* (*l-use*) to represent the reification of the processing of the utterance as an event/action conforming to some set of computational and lexical actions, specified for a particular "language", according to which metalinguistic judgments can be assessed.[22,23] Here the potential to introduce such language-use metavariables makes explicit the freely available potential for switching the language, idiolect, or any other variety of usage, and evaluation of metalinguistic judgments according to such switches, all of which can occur at any subsentential stage of production/interpretation.

The DS-TTR grammar operates by means of licensing in context word-utterance events according to their temporal order. As we said, words (and the operation of "syntax" in general) are modelled in DS-TTR as offering 'affordances', opportuni-

[21] In the illustration of such phenomena, in my view, scripted dialogue provides valuable evidence because such occurrences cannot easily be attributed to speech errors.

[22] Note that this also shows that the above mentioned CONTEXT vs. CONTENT distinction is indeed artificial and hence present here only for simplicity of display purposes. The truth values of "metalinguistic" statements rely on conceptualisations of the instantiation of implicit contextual variables.

[23] Metalinguistic judgements (involving concepts like 'sentence', 'word', etc.) involve **G**-dependent types that range over conceptualisations of NL-use that reflect folk-linguistic conceptions but do not necessarily correspond with the analysts' grammar of a particular language (unless of course the discourse involves discussion of exactly such a grammar). The actual processing model (the *grammar*) used (unconsciously) for processing an utterance will be captured by the rule-level variable indicated as ***g*** in the quotation-related processing actions later. Unlike G&C, this is an essential reservation for the DS-TTR formalism which does not license form-meaning pairs ("expressions") but, instead, interlocutors' performance, i.e., the production and interpretation of actions. Any reification of (part of) the products of such actions is then necessarily the outcome of some coercion and reification of the actual language use.

ties for action, exploited by the interlocutors to facilitate interaction, so that words and linguistic constructions are not conceptualised as abstract objects, 'expression types', that are associated with referential/semantic values (cf. Cappelen and Lepore 2007: chapter 12). As in DRT (Kamp 1981; Kamp and Reyle 1993) and related frameworks (see also Jaszczolt 2005), semantic, truth-conditional evaluation applies solely to contextually enriched conceptual representations. However, unlike all these other models, truth-conditional evaluation applies incrementally, as each word is processed (see e.g. Hough 2015 for details). The other distinguishing feature of DS-TTR, as compared to DRT, is that the process of progressive building of conceptual structures is the only notion of "syntax" admitted, in that there is no intermediate level of syntactic structuring where a string of words is assigned hierarchically organised constituency as syntactic categories, phrases or clauses. Such constituency is considered in DS-TTR as epiphenomenal on the function-argument semantic relations as typified in the lambda-calculus analyses of NL meanings. In consequence, in DS-TTR, all standardly assumed syntactic dependencies have been reformulated in procedural terms, i.e., in terms of how time-linear processing is affected by semantic dependencies. Such procedural explanations include, in particular, the classical data used to deny 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). With no privileged semantic entities corresponding to (types of) expressions, only mechanisms for processing stimuli, quotation thus offers a crucial test for the legitimacy of these DS-TTR claims regarding natural languages: When processing a quoted/cited string, what happens within the quotation marks (or any other indications) following these assumptions?

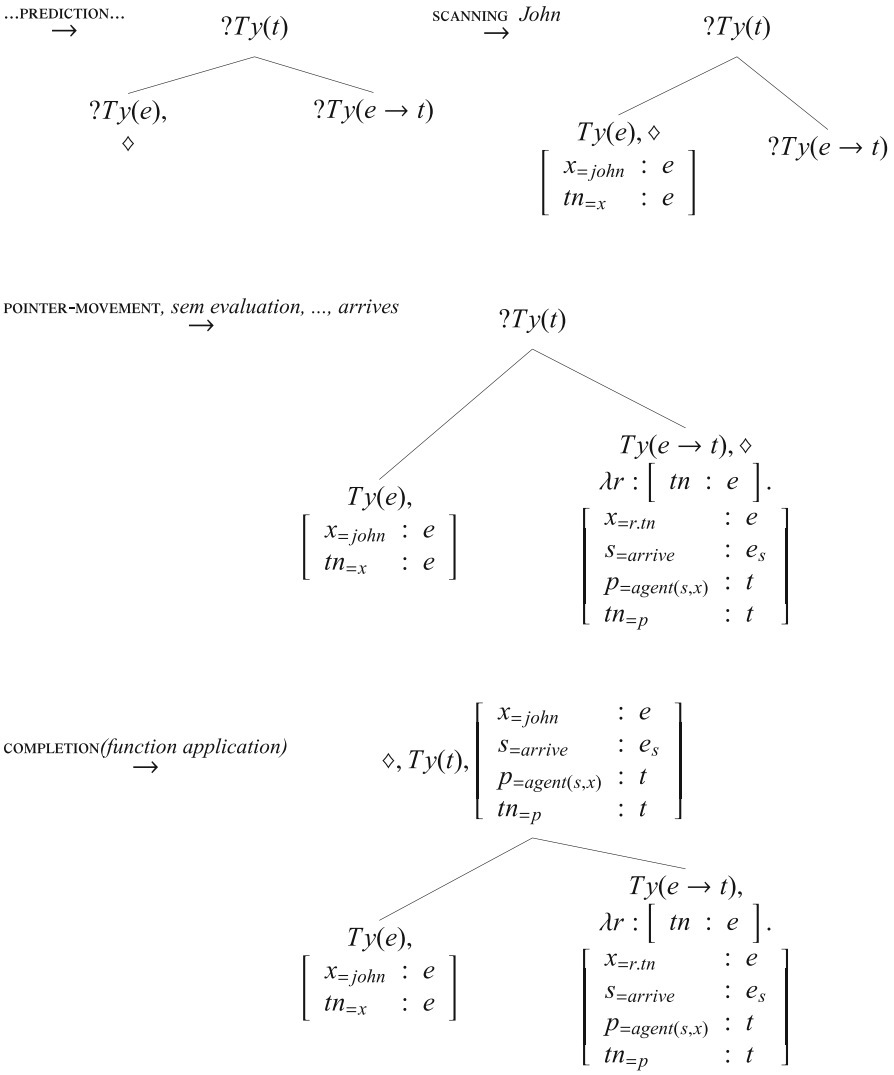
To answer the question of NL quotational/citational uses (Sects. 7.3.1, 7.3.2, and 7.3.3) we first need to remember that the application of these DS-TTR grammatical assumptions to the analysis of quotation is parallel to their application in the analysis of conversational mechanisms. This is because, as we saw in Sect. 2, and following the insights of Ginzburg and Cooper (2014), quotational phenomena appear to be subsumed under the constructs needed to underpin interactional mechanisms and the modelling of metacommunicative coordination. From this point of view, first, DS-TTR's lack of a syntactic level of representation and its sub-propositional semantic evaluation is an advantage in conversational modelling since it directly provides the mechanisms for accounting for split utterances and fragmentary discourse in dialogue (see (25) and the illustration in (85) below). Various cases of subsentential actions in dialogue are employed to indicate that the words uttered by the current speaker do not necessarily reflect his/her perspective (as in e.g. (11)–(17), (25)), or are not being used with the sole purpose of inducing their conceptual content (see e.g. (18)–(24)).

In DS-TTR, modelling the potential of partially assuming another interlocutor's role, being perceived as "demonstrating" what the other interlocutor was going to say, is achieved unproblematically because the potential for sharing syntactic/semantic dependencies is guaranteed at each step and there is no requirement to derive a global propositional speech act: Both speaker and addressee perform processing steps incrementally, guided not solely by the NL string, but also driven by *predictions* (introducing 'goals') generated by the DS-TTR grammar (in the displays these anticipated goals are shown with a *?* accompanying each predicted, but not yet realised, specification). 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 production and parsing start with the expectation of the appearance of a subject first (of semantic type e, *?Ty(e)*), followed by a predicate afterwards (of semantic type $e \rightarrow t$). The pointer then shifts to the *?Ty(e)* node, which processing of the first word in the sentence, e.g. *John*, annotates with a value of type entity (e.g. the logical representation of the individual John which is indeed of *Ty(e)*). Subsequently, if an intransitive verb, like *arrive*, follows, it will trigger actions that annotate this predicate node with a function to be applied to the subject. It will also introduce the event/situation (shown as the variable s below of type e_s) that is taken as the witness of the type derived by processing the clause (see earlier Sect. 4). Finally, computational actions that complete the process will follow next (CONTEXT values are omitted for clarity, the label *tn* indicates the treenode address which serves as a handle for accessing the relevant node content, PREDICTION and COMPLETION are examples of the general non-lexical computational actions employed in DS):

(84) Incremental steps in processing a clause with an intransitive verb[24,25]:

[24]The representations here employ so-called *manifest fields*. The notation employing the equality sign is abbreviatory for a singleton type constructor (see e.g. Cooper 2012; Ginzburg 2012), indicating subtypes of some type restricted to a single member, that is, only the relevant value mentioned. So, for example, $x_{=john} : e$ means that the value of label x is of the subtype of type e whose unique witness is the individual John.

[25]The notation employing a dot indicates a path to a value, e.g. *r.tn* indicates that the value needed is to be found as the value of label *tn* in record (type) r (see also fn. 24). (Note that this use of the dot notation is different from its use in separating the λ-bound variable [plus restrictions in TTR] from the function expression, e.g. $\lambda x{:}[x_{=john} : e].\ Arrive'x$)



If a transitive verb follows instead, its lexical entry will introduce not only the conceptual content associated with the verb but also the prediction that an argument, the complement, will follow immediately afterwards. Such complements can be either of individual entity type (type *e*) or of propositional type (semantic type *t*), the latter for propositional attitude or reporting verbs. The embedding of propositional types as complements defines one aspect of linguistic recursivity. Another aspect, related to adjunction (e.g. relative clauses, adverbials, parentheticals), is implemented by relating trees via a so-called LINK relation, a relation that does not involve mother-daughter tree-relations. The construction of a LINK relation among two independent trees offers opportunities for interrupting the construction of one

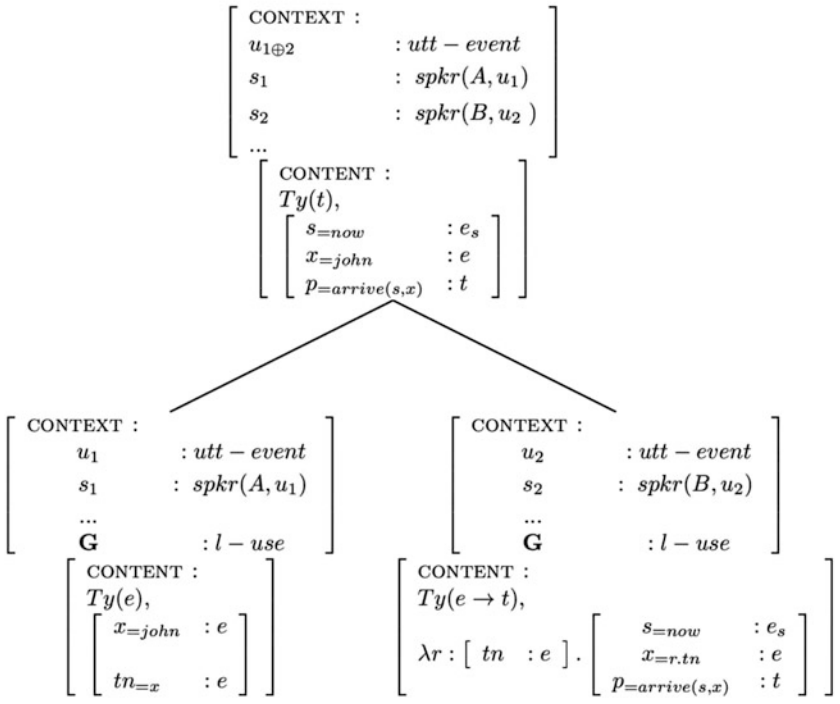
tree at a specific node in order to elaborate on some of its terms by shifting the pointer to an auxiliary tree, processing some linguistic input there, and, eventually, enforcing sharing of this information among the paired trees.

Thus, *parsing* in DS-TTR 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 on checking that the string thus produced can deliver a conceptual representation that accords with the (partial) conceptual structure the speaker attempts to verbalise (called the *goal tree*). As a result, speaker and hearer roles involve in part mirroring each other's actions (Gregoromichelaki et al. 2013a, b; Pickering and Garrod 2013). From this perspective, it is then unproblematic to model the sharing of utterances and the joint construction of conceptual structure in dialogue. As the schematic illustration in (85) shows, the only difference that registers the change of utterers during simple split utterances is the change of values in the contextual parameters:

(85) Processing *John arrives*: final content derived through two micro-conversational events by different speakers

A : John ...
B : arrives

$\mapsto$



The sharing of syntactic/semantic dependencies is possible because, as speakers and hearers simulate the actions of each other, the fulfilment of syntactic/semantic predictions is attempted at each incremental step, subsententially, for both parser (hearer) and generator (speaker). Such fulfilment can be achieved by either speaker or hearer, whether on the basis of the other interlocutor's input, the context, or by recourse to the processor's own resources. As no structure is ever assumed to be derived for the sentence string, no whole-string "grammaticality" considerations ever arise. Similarly, no context-independent compositionality restriction applies to NL strings; only contextually derived conceptual structures are interpreted compositionally. Hence, fragments that can be processed by fitting into a structure that is already in the context are licensed directly, that is, they are NOT characterised as elliptical and there is no requirement that they need to be enriched to a propositional type to be interpreted:

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

Such split utterances are unproblematically processable and are in fact a natural consequence of such a fine-grained bidirectional incremental system: As predictive goals are constantly generated by the grammar, to be achieved symmetrically by both the parser and the producer, the hearer/parser can wait 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 generation 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 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, 2012, 2015; Gargett et al. 2008, 2009; Gregoromichelaki et al. 2009, 2011, 2012, 2013a; Kempson et al. 2009a, b, 2011a, 2012; Purver et al. 2009, 2010, 2011).

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

7.3 Metalinguistic Devices in DS-TTR

7.3.1 Pure Quotation (Citation)

As we've already seen, the utterance-situation parameters (speaker, hearer, time of utterance, etc.) in the information state, the value of CONTEXT in DS-TTR, include storage of the word forms that have triggered processing. As Ginzburg (2012) has shown, this is essential for various parallelism effects observed in the processing of dialogue phenomena (e.g. interpretation of clarifications as echoic; see e.g. (5)). In addition, CONTEXT also stores the DS-TTR processing actions that have already

been used in deriving conceptual CONTENT structures. This is necessary for the resolution of anaphora and ellipsis (Kempson et al. 2001, 2011b, 2016; Cann et al 2005, 2007; Gregoromichelaki et al. 2011; Kempson et al. to appear; Purver et al. 2006). Under this view, the processing actions utilised in parsing and production are first-class citizens in the model in that the grammar includes means for referring to sequences of actions already stored in the CONTEXT, reasoning over them, and reemploying them again in subsequent steps (Cann et al. 2007). This is necessary for the explanation of phenomena like 'paycheck anaphora' and 'sloppy readings' of ellipsis ((87)–(89) below) where the interpretation changes due to the new local environment where the anaphoric elements acquire their interpretation (Kempson et al. 2011a). They also need to be available both subsententially and for anaphoric and cataphoric employment, the latter shown in (90)–(91) below:

(87) The man$_i$ who gave **his paycheck** to his wife was wiser than the man$_j$ who gave **it** to his mistress.
['man$_j$'s paycheck']

(88) John upset **his** mother. Harry too. ['Harry upset Harry's mother']

(89) The man who arrested John failed to read **him his** rights. The man who arrested Tom did too. ['failed to read Tom's rights']

(90) The representations here employ **so**-called *manifest fields*. [this document, footnote 23]

(91) **It** appears that John left.

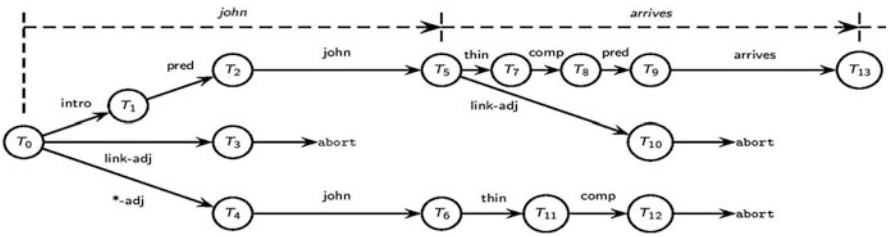
In cases like (87)–(89), in order to model the rebinding of the anaphoric elements (indicated in bold) to the newly introduced subjects in the next clause, DS-TTR retrieves a sequence of actions <a_i, ..., a_{i+n}> already performed in processing the previous clause and therefore stored in the context representation. It then executes them again in the new sentential environment with the result that the new subject now provides the local binder of the metavariables introduced by the anaphoric elements (see e.g. Purver et al. 2006; Cann et al. 2007).

The same action retrieval mechanisms are used in cases of *self-repair* where one element that replaces another (*surfing* to replace *swimming* below) needs to re-compose with elements that have already been processed (*with Susan* below):

(92) Peter went swimming with Susan, um . . . , or rather, surfing, yesterday. ['Peter went surfing with Susan yesterday']

Such cases of *repair* (whether self- or other-repair, including clarification), in many cases, require re-execution of already processed material. This is modelled as the re-running of a sequence <a_i, ..., a_{i+n}> of actions stored in CONTEXT in order for material to be reprocessed (Hough 2015; Eshghi et al. 2015). For such repair and other purposes, DS-TTR also records the various potential but not pursued processing options licensed by the grammar at each step: As we saw, the DS-TTR formalism operates by generating predictions regarding future steps of processing. This results in the generation of multiple potential processing paths that the information state can develop into. These paths, whether pursued or not, are taken as part of the context representation and can be illustrated in the form of a DAG, a Directed Acyclic Graph (see Sato 2011; Hough 2015 for formal details):

(93) Context DAG showing various potential processing paths



In such a graph, edges correspond to potential DS-TTR computational/lexical actions and nodes to resulting information states (Purver et al. 2011). This contextual representation is employed for the modelling of various dialogue phenomena, where the parser needs to backtrack to a previous path, other than the one actually pursued, and proceed to another interpretation of the input or reformulation of the utterance (see Hough 2015; Eshghi et al. 2015 for details). The claim that such abandoned paths need to remain in the context is additionally justified by cases where "repaired" elements need to be accessible for e.g. anaphoric purposes:

(94) **Jill** left, no, (I mean) Bill left, **she**'s in Paris already.

Now let's examine the cases of Recanati's "closed quotation" where an NL-string appears in a regular 'NP' position, i.e., where the grammar, under DS-TTR assumptions, has already generated a prediction for the processing of a singular term (or any other semantic type).[26] Given this prediction, there will always be an attempt for whatever is processed in such a position to be construed as 'subject', 'object', modifier etc. Exactly because any DS-TTR grammar for a particular use of language consists of routinised sequences of actions, this will also be possible for any set of actions, for example non-linguistic actions as in (73)–(75). Moreover, given incrementality and the absence of sentential grammaticality licensing, any DS-TTR model can license the processing of input provided through some language use distinct from the one providing the tree position that the content of this input will annotate.

Regarding interpretation, as word forms in DS-TTR are assumed to constitute triggers for macros of actions, which include importing conceptual content contributions, inevitably any conceptual contribution associated with a cited word or string of words will become available to the interpreter if it belongs to a known type of language use; and the same goes for any other conventionalised non-linguistic signals. However, where the context requires a "metalinguistic" interpretation for the uttered string, the conceptual value, like other properties of the stimulus associated with the word-form, even though accessed and built up, ends

[26] For the potential of such quoted strings to function as Ns or other categories rather than NPs (e.g. *The whys raised by this issue*. *These are not 'I really should' radishes* . . . . (Clark and Gerrig 1990, from Jon Carroll, *San Francisco Chronicle*)), see De Brabanter (2005b, 2013).

up embedded as the value of the particular predicted type on the treenode of the eventual conceptual representation. In such cases, given that the DS-TTR grammar does not provide form-meaning correspondences but only provides for the parsing and generation of utterance events, the process of 'coercing' some linguistic element to fill the role of an already predicted conceptual type on the tree can be taken as a reification of the grammatical process itself (like the demonstration of a car sound in the same position can be taken as the conceptualisation of some sound experience that is being demonstrated).

In DS-TTR terms, these assumptions can be expressed as the ad hoc categorisation of the running of a sequence of actions $<a_i, ..., a_{i+n}>$ at a node. Such a sequence will belong to some particular linguistic use (grammar) indicated as the rule-level variable ***g*** (see earlier Sect. 7.2 and fn. 19, 21, 22) which becomes instantiated by the intended grammar being invoked. The idea is that embedding the actual execution of a sequence of actions as the conceptual value of a node on the tree results in their conceptualisation as an element of the type already predicted in the particular tree position where the pointer finds itself. The intuition behind this implementation is that an utterance event (notated as u_q below) is performed (demonstrated) under the assumption of a particular DS-TTR grammar g (captured by binding of ***g***) in order to provide the content value for the current treenode[27]:

(95) Computational action for processing quotation:

IF $?Ty(\mathbf{x}_{\in\{e,cn,\ldots\}})$

THEN put $Ty(\mathbf{x})$

put $(u_{q=\mathbf{run_g}(\langle \mathbf{a}_i,\ldots,\mathbf{a}_{i+n}\rangle)}: e_s)$

ELSE abort

The **IF** condition in the computational macro first checks whether the pointer is at a node predicted to be of a particular type, e.g. type *e*, *cn* (common noun), or any other type of content that cited strings can be associated with ($\mathbf{x}_{\in\{e,\ cn,\ \ldots\}}$). If it can be shown, as seems to be the case, that content derived from citation can belong to any semantic type (De Brabanter 2005a, 2013), this restriction ($\in\{e, cn, \ldots\}$) can be dropped. Suppose that the pointer is at a node predicted to be of type *e*. The string to be processed is the following:

(96) 'John arrives' is grammatical.

In such a case the action in (95) can apply to provide a value on this node by processing the upcoming utterance event u_q which is immediately provided (note that events in DS-TTR belong to subtypes of type *e*, the type e_s (Gregoromichelaki 2006; for the notation employing the = sign of manifest fields employed here, see fn. 23). Unlike the anaphoric cases of action re-running we saw earlier in (87)–(89), here the process is *cataphoric*: the demonstrating event has been predicted to occur

[27]Like $\mathbf{x}$, $\mathbf{a}_i$,..., $\mathbf{a}_{i+n}$ are also rule-level variables that become bound to whatever individual actions the current state provides; see fn. 19.

and its subsequent occurrence and processing duly satisfies this prediction. In this way, the content that appears at the relevant node, the specific singleton type of the event u_q, its grammatical characterisation in the DS-TTR sense, is constructed on the fly: this type of event has as its sole witness an upcoming employment of some grammar ($\texttt{run}_{\mathbf{g}}$) to execute a sequence of actions <a_i,..., a_{i+n}>. Notice also that the instantiation of the metavariable indicated by the parameter ***g*** will invoke a language use that can be distinct from the linguistic use applicable to the rest of the string. The folk-linguistic (or scientific) characterisations that predicate of such reified linguistic uses can then target aspects of the grammatical processing that has just been executed. This, in essence, is just a process of *explicit* categorisation of various aspects of the stimulus (for the potential properties accessed in such cases, see e.g. Saka 2011, 2013; for inferences narrowing down such targets, see e.g. Saka 2003/2005).

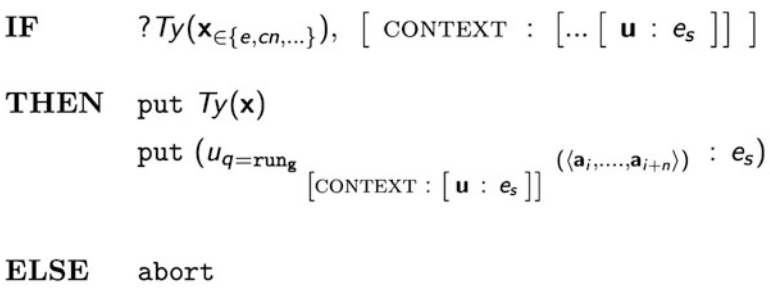
In (the unusual) cases where the metalinguistic/metacommunicative interpretation becomes available only after a metalinguistic predicate has been processed, the parser will need to backtrack along the DAG-recorded path (see (93)) to a previous parse state in order to pursue this new option as in cases of repair in conversation. We can then assume that the function of quotation marks or other quotational indications in spoken or written language is exactly to indicate to the parser that a non-default processing strategy (i.e. a DAG path of low-probability rating) is to be pursued.

Such cases are similar to those where there is invocation of a previous utterance event ('echoing'). For example, in a situation where A rehearses the string *John arrives* for a radio play, B, the sound engineer, can say:

(97) "John" was a bit loud.

In such cases, we assume that, in addition to the choice of grammar *g*, there is also a contextually available token utterance event (in the display below indicated by the rule-level variable ***u*** belonging to the subtype of entities that constitute events/situations e_s). As in the previous case (95), there is again a quotation event (u_q) mentioned below, conforming to some linguistic use ***g*** (instantiated by *g*), but this time (part of) the contextual parameters are set by the contextually available event (instantiating ***u***):

(98) Computational action for processing sententially embedded echoic quotation:



These non-sentential echoing quotation cases are similar to direct quotation to which we now turn.

7.3.2 Direct and Indirect Reports

Under DS-TTR, the lexical action for a framing verb (e.g. a *verbum dicendi* like *say*) can be assumed to uniformly provide conceptual content that is able to combine with related semantic objects (e.g. propositions, of type *t*, and utterance events of type e_s), provided through distinct combinations of processing actions. Such combinations of processing strategies can result in cumulative results modelling either direct and indirect reports, or intermediate phenomena, without postulating specific types of static syntactic constructions as in G&C and other models. In DS-TTR, the only factor that accounts for the alleged syntactic differences between direct and indirect quotation (e.g. parenthetical/reversed word order or *wh*-extraction; see e.g. Schlenker 2011; Bonami and Godard 2008) is that the actions induced by such verbs, like other verbs in English (e.g. *eat*) can license 'object-drop', a license that is modelled in the DS-TTR account by allowing object-drop verbs to take as their complement a *metavariable*. As in the DS-TTR modelling of ordinary cases of pronominal or elliptical anaphora resolution, such a metavariable needs to be provided with a value from CONTEXT (in the form of existing conceptual content or via the rerunning of actions). In 'direct quotation' cases, the value for such a metavariable can be provided by the independent clause processed as an antecedent either anaphorically or cataphorically, e.g.:

(99) "I talk better English than the both of youse!" John shouted/announced/said.

Framing verbs can also compose directly with non-linguistic actions, e.g. gestures or sounds, which is straightforwardly modelled in the DS-TTR formalism, as there is no distinction between linguistic and other actions: both invoke categorisation processes which, in the case of linguistic performance, are standardly characterised as the 'grammar'. However, as has been pointed out previously (Slama-Cazacu 1976; Clark 1996; Postal 2004; De Brabanter 2010b), the grammar needs to be conceived in a much wider sense to account for cases like the following:

(100) The car engine went [brmbrm], and we were off. (Clark and Gerrig 1990)
(101) The boy who had scratched her Rolls Royce went [RUDE GESTURE WITH HAND] and ran away. (Recanati 2010)
(102) I didn't see the [IMITATION OF FRIGHTENING GRUMPINESS] woman today; will she be back this week? (De Brabanter 2010b)
(103) Piano teacher to student: It's not [plays passage in manner μ]—it's [plays same passage in manner μ']. (Horn 1989)

As the DS grammar operates predictively, positions in the conceptual representation are constantly generated in anticipation of the next input.[28] Being processed in such a position can coerce any perceptual stimulus to induce a processing action that will compose its derived categorisation (i.e. content) with the rest of the conceptual representation (Gregoromichelaki 2013b). Since such conceptual representations are expressed through the TTR formalism in DS-TTR, as in G&C, any perceptual stimulus can be accommodated by the type system via the subtyping relation (see Cooper 2012 for formal details), hence allowing for the construction of ad hoc types. This is commonplace in actual conversational interactions. For example, Gregoromichelaki (2012) and Gregoromichelaki and Kempson (2013) argue that non-linguistic actions are regularly the antecedents of ellipsis, clarification requests, etc. In such cases, constraints on the conceptualisations of such actions are imposed linguistically via the form of antecedent-requiring elements (e.g. case requirements[29]; see (104) below) that retrospectively restrict the structure of the construal underpinning the conceptual representation:

(104) [Context: A is contemplating the space under the mirror while re-arranging the furniture and B brings her a chair]
A to B:tin karekla tis mamas? / *i karekla tis mamas? Ise treli? [clarification] [Greek]
'the$_{\text{ACC}}$ chair$_{\text{ACC}}$ of mum's / *the$_{\text{NOM}}$ chair$_{\text{NOM}}$ of mum's.
Are you crazy?'
'Mum's chair? Are you crazy?'

(105) [Context: A asks who C has invited and D points to B]
C: (Actually,) not him, his sister.

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

Supporting DS-TTR's assumed uniformity of lexical, computational and context-shaping non-linguistic actions, notice that even the sequential process of parsing/production can become the object of anaphoric mention:

(107) The rules of Clouting and Dragoff apply, in **that** order. (Ross 1970)

The same idea covers cases of direct and indirect quotation: the fact that contents supplied by framing verbs can acquire propositional complements either directly via embedding a description of the content of an utterance event (indirect reports) or indirectly via the echoing of a previous utterance event (propositional or not, in direct reports) allows us to capture the continuity of direct/indirect discourse. It also explains the intermediate cases, e.g. mixed quotation and free (in)direct

[28] As we saw, the position currently under development is indicated by a 'pointer', ◇, which is what accounts for variable word-orders.

[29] According to DS-TTR, case affixes in morphologically rich languages impose the prediction/goal of an appropriate tree-structural position to accommodate the conceptual content contributed by the linguistic element carrying it.

discourse structures. Such structures show that fine-grained processing mechanisms can be combined in various ways, both synchronically, at the discretion of a current speaker for novelty effects, and diachronically, becoming routinised and therefore commonplace (i.e. assigned high probability as processing paths in the context DAG; see earlier (93)), to deliver various conceptually articulated construals and non-conceptual impressions. Thus modelling these intermediate phenomena via fine-grained mechanisms that can combine with each other argues against postulating monolithic, fixed form-meaning correspondences ('constructions') since the available mechanisms can, and will, be freely exploited by human processors to deliver various novel effects in context. Fixing ab initio the outcome of such combinations is bound to fail to account for the various potential outcomes of situated processing (see also Gregoromichelaki and Kempson 2017).

Turning to direct quotation first, we can describe the DS-TTR grammatical mechanisms allowing for its processing and effects as follows: As the DS-TTR grammar is articulated in terms of actions, we can postulate that the properties that characterise "direct reports" are the result of potentially choosing to focus the hearer's attention[30] on the triggers of the lexical actions (*words*, as stored in the context DAG) presented as having been used by another speaker (i.e. demonstrating (echoing) a contextually available utterance event $\boldsymbol{u}$ retrieved from and stored in the CONTEXT representation and via instantiation of the $\boldsymbol{g}$ parameter to another speaker's grammar as we saw earlier for pure quotation). We can assume that this can sometimes be indicated by the quotation marks. As discussed earlier for pure quotation (Sects. 6 and 7.3.1), following G&C, Recanati (2010), and Predelli (2003), for direct discourse the DS-TTR conceptual representation derived will involve an embedded utterance event u_q, corresponding to the demonstration the speaker performs. The verb *say* in English and other languages regularly combines with utterance events, whether echoic or not or assertional or not (contra Brandom 1994: 531 whose presentation implies that the propositional-complement use should be primary in that it makes explicit implicit 'assertional ascription' practices):

(108) He said "constraints in agriculture" when he meant "excluded products".
(109) At 36 months, he had begun developing functional language but could not grasp concepts like first and second person; he said "I" when he meant "you," and "you" when he meant "I." [https://findingmykid.com/tag/hockey/ retrieved 13/6/16]
(110) Martin approached, listened for a second, looked at the UPS guy, and said, "Oh, hi!" [https://findingmykid.com/tag/hockey/ retrieved 13/6/16]
(111) "Ouch!"/"Wow", she said.
(112) Say "Kognitionswissenschaft".

[30]When "focussing attention" is explicitly conceptualised, it is represented in DS-TTR by introducing additional inferentially derived propositional or sub-propositional contents as LINKed structures, the representational device used for the processing of adjuncts like relative clauses, conditionals etc. (see Kempson et al. 2001; Cann et al. 2005; Gregoromichelaki 2006).

In cases of direct reports, this demonstrating event u_q, "echoes", i.e., bears a contextually determined similarity value, 'resemblance' to, another, anaphorically retrieved utterance event (***u***). This anaphorically retrieved event supplies (part of) the contextual parameters, thus accounting for the corresponding change in the values of indexicals across speakers and turns:

(113) A: Will **you** say to Nick. . .
B: "I hate **you**"? Yes, why? ['B hates Nick']
(114) A: Did **you** say to Nick "**You** . . .
B: "hate **yourself**"? Yes, why? ['B said Nick hates Nick']

Notice that, as in the earlier (104)–(107) cases, the presumed contextually available element, in this case a speech event, need not be part of the context already; instead, the introduced requirement that it should be part of the context eliminates DAG paths where it is not possible for such an event to be conceptualised (e.g. contexts where the hearer believes that the reportee was unable to communicate) or leads to the generation of further metacommunicative interaction, e.g. clarification, in order to be accommodated. This is standard for many cases of direct reports where what is "reported" has never actually been uttered (see e.g. Tannen 1986; Norrick 2015) and cases intermediate between direct reports and free direct speech where the contextual parameters again need to recruited from such an imaginary, reconstructed event:

(115) Adam: Well. I can tell you what her view on that is. and that
Sherm: what.
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] (Grimshaw 1987)

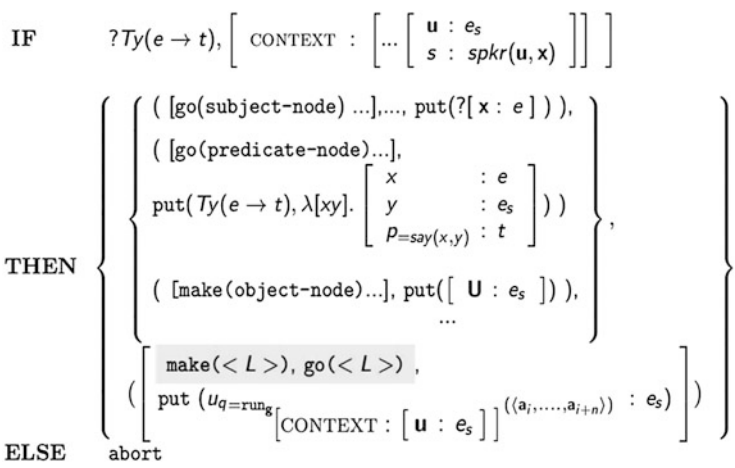
For cases standardly regarded as clear-cut cases of direct report constructions, in DS-TTR terms, the only difference with the previous echoing case in (97)–(98) is that such an otherwise freely available computational action has been "lexicalised": it has become part of the routinised macro stored as an option in the lexical entry of the verbum dicendi,[31] so that, in terms of the DAG representation, its execution constitutes a highly probable option. So, for example to process a string like the following:

(116) John said "I was loud".

the following (schematic as regards irrelevant details) lexical entry for *say* can be invoked:

[31] Alternative options in a lexical entry are listed as embedded in **ELSE** statements, before `abort` is encountered.

(117) Lexical entry for *say* + LINK sequence[32]:



The condition **IF** here expects the presence of a salient utterance event in the context (to bind ***u***) whose speaker will provide the value for ***x***. The lexical macro then ensures that the subject of the proposition will be that speaker ***x*** (put(?[**x** : *e*])). Next it constructs the predicate and its object node (abbreviated presentation here, see fn. 31) and inserts a metavariable ***U*** of type utterance event (e_s) as a temporary place-holder. A LINKed node is then introduced (shaded in the display), which is the device used in DS-TTR for the processing of adjunction (see e.g. Kempson et al. 2001; Cann et al. 2005). The conceptual value on this LINKed node will be provided by the execution of the actions needed to process the following string with contextual parameters provided by the contextually instantiated value of ***u*** which ensures that the indexicals receive appropriate values, e.g. as instantiated by the utterer of ***u*** for a pronoun like *I*. The DS-TTR constraints governing LINK transitions will then ensure that the value of the metavariable ***U*** will be unified with the content of the LINKed node.[33]

Under this analysis, the syntactic/semantic opacity observed in such structures is explained by (a) the presence of the LINKed node, as is usual in DS-TTR regarding the modelling of the banning of extraction from adjuncts (Kempson et al. 2001), and (b) the embedded nature of the propositional content derived on the LINKed node, in that it is just one of the TTR types characterising the utterance event.

[32] As mentioned earlier, `make`, `go`, `put`, `run`, etc. are elementary DS actions processing strings and building conceptual structure. They are modelled via accessibility relations among information states in the Dynamic Logic underpinning DS (see Kempson et al. 2001: chapter 9; Cann et al. 2007) The specifications `object/subject/predicate-node` are just schematic name abbreviations to avoid the clutter of presenting actual DS-TTR step-by-step actions and modalities.

[33] Some collections of sequences of actions are indicated as freely ordered or optional through bracketing to account for variable word-orders.

However, given that even in such cases the conceptual representation contributed by the quoted string is inevitably derived, 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:

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

(119) "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. (Maier 2017, to appear)

(120) A: I talk better English than the both of youse!
B: You obviously don't. (Partee 1973)

Instead of assuming that the availability of such anaphoric resolutions is the result of presuppositional elements or implicatures (as in Maier 2014a), here the grammar itself provides the resources for explaining the phenomena. As stated earlier, the resolution of both ellipsis and pronominal anaphora in DS-TRR is assumed to involve reuse of terms annotating CONTENT fields on treenodes, non-linguistically provided content, or the rerunning of processing actions stored in the CONTEXT (Eshghi et al. 2012; Kempson et al. 2015). 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 conceptual representation, they are stored in the context DAG and are available to be invoked for the resolution of anaphoric and elliptical occurrences as in (118)–(120). For the same reason, as in the G&C analysis, we can account for cases of "mixed predication" where both token and type aspects are addressed simultaneously; recall (38)–(39), repeated here:

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

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

But further than any other account, the present analysis extends to cases where the continuation of an utterance started by an initial speaker without any quotational intent can become quotational, i.e., treated as a demonstrating event (123)–(124); and conversely, structures initiated without an already present reported event which can be provided a quotational, echoing complement by the actions of another speaker (125):

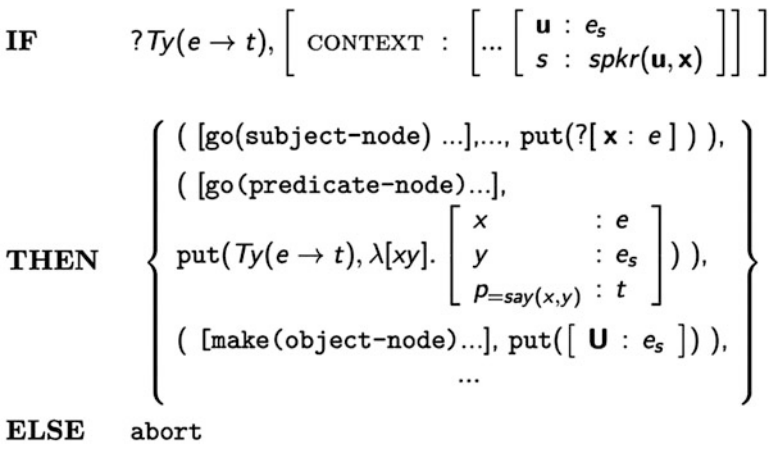
(123) Jem: Mary, whatever it is you think you know you mustn't speak of it. Not if you want to stay safe.
Mary: says the horse-thief [BBC Transcripts, *Jamaica Inn*, Episode 1]

(124) Miriam: That is the nastiest, dirtiest thing anyone has ever done
Patience: says Black Peter's strumpet! What are you crying for? [*Jamaica Inn*, Episode 1]

(125) Noel: What I'm saying is
Stacey: you are IT!
Noel (ironically): Well, yeah...
(adapted from BBC Transcripts, *Never Mind the Buzzcocks*, 16/10/11)

In cases like (123)–(124), there is anaphoric use of the reported utterance event (that is, the demonstrating event has already been performed instead of being executed after it has been announced as the quotational cataphoric uses we have analysed so far). In accordance with (117) earlier, the actions induced by *say* are executed but the value for the metavariable ***U***, the quoted event, is provided by appropriation of the other speaker's utterance that has just occurred, instead of being provided via an independent demonstration on a LINKed node. This utterance also constrains the value of the upcoming subject via the predicted unification with the speaker (*spkr*) value of the reported event available in the CONTEXT part of the **IF** condition (see also fn 32; this LINK-unenriched option also accounts for further variable [and parenthetical] word-order patterns in direct reporting structures)[34]:

(126) Lexical entry for *say* + direct report:



On the other hand, (125), repeated below, is an intermediate case of indirect report in English where the complementiser *that* is missing, as shown by the intended values of the indexicals[35]:

(127) Noel: What **I**'m saying is
Stacey: **you** are IT! ['Noel is IT']
Noel (ironically): Well, yeah...

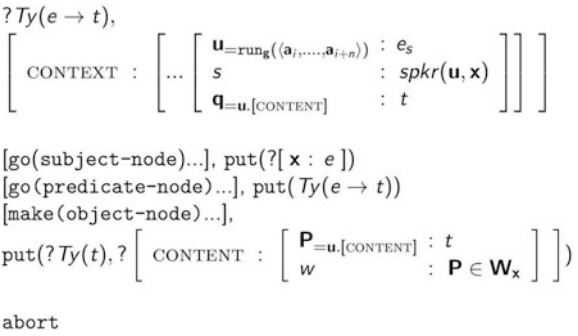
In such cases, the object node of the verb *saying* will be provided a value of type *t* (*Ty(t)*), i.e. the type 'propositional', which in DS-TTR, is a complex record type whose *tn* value, the final type derived, is *t* and does not carry any assertional

[34]Note that due to the implementation of incremental licensing, parsing/generation in DS-TTR can be initiated from any subpropositional stage, e.g. here starting with the requirement to build a predicate (*?Ty(e → t)*).

[35]Complementisers in DS-TTR do not themselves contribute content that appears on treenodes, they just execute procedural functions of introducing constraints on what can occupy nodes or predictions of upcoming input.

implications (Gregoromichelaki 2006). Accordingly, this is what explains the syntactic transparency of such structures (e.g. extraction possibilities; see earlier (69)–(70)), as is usual in DS (see e.g. Kempson et al. 2001; Cann et al. 2005). This will be ensured by means of the lexical entry for *say* that combines its content with a propositional complement as shown below.

(128) Lexical entry for *say* + indirect report:



Simplifying the semantics for illustration purposes (see also Maier 2017 for a similar formal implementation), here the situation derived as part of the content of the embedded report (the witness of the proposition, the value to replace the metavariable **P**) is constrained to exemplify the same type as the propositional content (**u**.[CONTENT]) of some uttering event (**u**) by the contextually invoked speaker (**x**, also the subject of the sentence) and to be part of all the worlds in the set of worlds compatible with what this speaker said (thus instantiating the value of the metavariable $\boldsymbol{W_x}$).[36] However, unlike direct reports, this uttering event (**u**) does not provide the contextual parameters for the report since the reporter's utterance just provides an interpretation of that uttering event. Notice though that, just like direct reports, such structures can also felicitously embed (descriptions of) conversational phenomena, e.g. repetition and abandoned sentential strings, which renders essential their incremental licensing:

(129) I kept up, and anxious not to lose him, I said hurriedly that I couldn't think of leaving him under a false impression of my-of my-I stammered. The stupidity of the phrase appalled me while I was trying to finish it, ... (Clark and Gerrig 1990, from Joseph Conrad, *Lord Jim*)

Now returning to (125)/(127), processing the continuation accompanied with context shift is unproblematic because each word micro-conversational event will introduce its own contextual parameters, hence accounting for the resolution of both the contents of *I* and *you* to same individual. The result will be a proposition ('Noel is "it"') that matches the hypothesised utterance produced by Noel even

[36]Further similarity requirements could be introduced following G&C's definition of similarity relations; the complications mentioned by Cappelen and Lepore (1997), regarding *similarity* of content rather than replication of contents, could be implemented by loosening the same-type restriction through appealing to the subtyping relation.

though this utterance has been produced by Stacey describing what Noel would have said (the pragmatic effect being that Stacey only "pretends" that this is the utterance that Noel would have produced, so, unlike genuine continuations, she carries the responsibility for its content so Noel has to confirm it). Since, on the surface, only content is relevant here, Stacey has to switch indexical when assuming Noel's speakership (see (83); cf. (31) and (113)). In modelling the processing of this structure, we assume that *what*, which is taken as an anaphoric element in DS-TTR, has introduced a metavariable for an event to be resolved cataphorically (for other such grammaticalised cataphoric structures, see (91) earlier and Cann et al. 2005; Gregoromichelaki 2013a). This metavariable will provide the temporary place-holder for the binding of the rule-level variable ***u*** in the lexical action in (128). This metavariable can eventually be provided with a proper value only after the second speaker, Stacey, has uttered her part with subsequent appropriate resolution of all remaining variables.

What allows the flexibility of such an account is the difference between this approach and G&C's, namely the fact that a monolithic utterance event is not necessarily derived at once for the whole complement of the framing verb. Instead, as the contextual parameters are reset at each micro-conversational event, there is the possibility at each subsentential stage for the current speaker/hearer to switch. For the same reason, the incrementality of DS-TTR also provides for the modelling of the potential a speaker has, even during a non-shared utterance, to be able to shift the default context and perform a demonstration. This is what accounts for both cases of free (in)direct reports and mixed quotation as we are going to see in the next section. In line with Recanati (2010), we can assume that standard uniform, non-shared indirect reports are cases where the CONTEXT field values remain constant throughout the utterance of both the reporting section of the sentence and the reported-content part. As a consequence, indexical elements receive their interpretations from the context established by the current utterance event U_n. However, as a consequence of the lexical action introduced by the framing verb, a new possible world/time (or set of world/times) metavariable ***W*** is introduced for the report to express the fact that it reflects the reportee's view (see also Recanati 2000). Such contextual and world/time parameters can be shifted independently of each other, and the possibility of shifting world and context parameters (including time of utterance) independently and incrementally, for each word-utterance (each *micro-conversational* event; see earlier Sects. 3 and 7.22) as the utterance develops, models the otherwise puzzling cooccurrences of transposed and untransposed indexicals considered by Recanati (2000: chapters 15–16) and pronouns and tenses in various intermediate cases of reporting (Eckardt 2014). Confirming the desirability of such flexibility, notice the independently established fact (Gregoromichelaki et al. 2011) that in cases of split reporting utterances, indexicals will acquire values according to who currently assumes the relevant interlocutor roles (see also (31)):

(130) A: So **you** say **you** will live
B: by **my** pen, yes

(131) A: Did **you** say to Nick that ...
B: **you** injured **me**? Yes, why? ('A injured B')

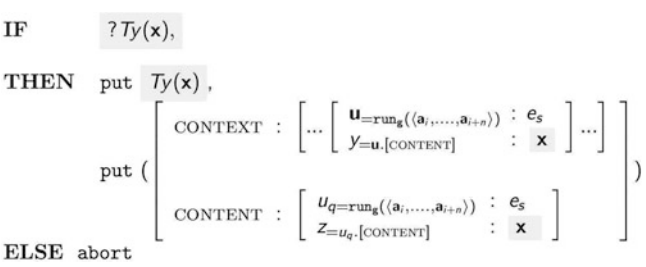
(132) A: Did **you** say to Nick that **you** ...
B: injured **myself**? Yes, my doctor says so. ('B injured B')

As we said earlier in Sect. 7.1, 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_1,$u_2 \ldots u_n$ subevents that resulted in a (perhaps joint) utterance-event $U = u_1 \oplus u_2 \oplus \ldots u_n$. Hence the interpretation derived eventually has the values of the indexicals as intended by the participants at each previous processing stage in that their lexical actions have been executed subsententially in line with 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 (132) fully licensed as joint utterances and provided with appropriate interpretations. Any other grammar that insists on an independent syntactic analysis of such strings (e.g. Potts 2007; Maier 2014a) will have trouble with such utterances, as the string of words *Did you say to Nick that you injured myself* will have to be characterised as ungrammatical (and for (130)–(131) it will derive the wrong interpretation).

7.3.3 Free (In)direct Discourse, Mixed Quotation and Scare Quoting

Essentially, along with Maier (2014a), the continuity of pure quotation, direct reports and mixed quotation is also assumed here; however, in line with G&C, the grammar does not need to implement this insight by employing special devices. Unlike G&C, since DS-TTR does not impose a separate level of syntactic analysis for the string of words, only the conceptual representation derived by processing the string, there is no issue arising here in terms of characterising distinct syntactic categories for indirect, direct, free, and mixed quotation structures in contrast to any other grammatical analysis of quotation (also, in fact, contra Recanati 2000, 2010). The only mechanism that is needed is the general mechanism in (95) that deals with pure quotation cases potentially accompanied with the assumption that there is an echoed event (utterance or thought):

(133) Echoing version of computational action for processing pure quotation with derived content:



The difference between the macro in (133) and the one in (95) is that in (133) the content derived by processing the demonstrating event (u_q), under a grammar instantiating ***g*** potentially distinct from the current speaker's grammar, and the type of content derived by the echoed event (the instantiation of ***u***) need to match (as shown by shading). That derived content will occupy the current node under processing, which can be of any type. Additionally, as an option, the contextual parameters can be provided by the echoed event as in the intermediate echoic case in (98) and the direct report cases in (117).

In consequence, to extend the coverage of the insights of the G&C account, DS-TTR does not need to employ specific constructions to deal with separate quotational phenomena, only mechanisms that can apply freely, combine with each other, and interact with the context, while at the same time eschewing a syntactic level of representation and definitions of abstract 'expressions' and 'expression types'. We now turn to the various remaining phenomena to exemplify briefly these mechanisms in various combinations. In the case of *free indirect discourse*, in addition to the free non-lexicalised introduction of an echoing demonstrating event, with or without shift of grammar, there is also a (non-lexicalised) shift in the CONTENT world parameter (as in the lexicalised option in (128); see Recanati 2000), for example the event is taking place in a world/time index according to somebody's thoughts/beliefs (hence this view reconciles the Maier 2014b and Eckardt 2014 analyses):

(134) Mary felt relieved. If Peter came tomorrow, she would be saved. (Recanati 2000)

Since in DS-TTR these parameters are independent, there is the possibility for independent shifting of world/time and CONTEXT parameters as required by particular linguistic elements and the discourse context (for systematising the grammatical constraints in this area, see e.g. Eckardt 2014). In DS-TTR, 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 echoing demonstrations that the speaker performs. Due to this fine-grained incrementality, there is no problem with having to coordinate the world/time and context shifts. This account gives results similar to those of Maier (2017, to appear)

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 independently the basis of the DS-TTR model. And, unlike other grammatical analyses, e.g. G&C and Sharvit (2008), since there is no independent level of syntactic analysis for the sentence, we do not have to license a complete sentential string that has to be internally consistent as to indirect/direct report features and contextual parameters (since, at the final stage, DS-TTR composes contents and not Kaplanian "characters"; cf. Eckardt 2014). Accordingly, *free direct discourse* (see (45) in Sect. 6) is simply a case where the CONTEXT parameters are also shifted uniformly along with the world parameter.

In the cases of *mixed quotation* (seen earlier in (49) in Sect. 6) and hybrid cases, there is no assumption here of any "verbatim requirement" (cf. Maier 2014a), so no such difference with indirect discourse ensues. Additionally, 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. It might also be the case that nobody has in fact uttered those words (hence scare quoting is not a separate phenomenon). Such cases can be adequately dealt with through the processing macros either in (95) or in (133):

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

(136) These are not "I really should" radishes . . . . (Clark and Gerrig 1990)

(137) Dutch is a "that I him have helped" language. (Abbott 2005, from Philippe De Brabanter)

We can also account for any "syntactic" binding effects in mixed quotation since even in structures licensed through the lexical entry for verbs with an indirect report complement, as shown in (128), the speaker, by employing in addition the actions in (95) or (133) for part of the utterance, can freely shift in and out of a demonstration:

(138) John said that "the queen of each man's heart" loves only herself. (Johnson 2011)

(139) Which houses did the FBI say they could "search without warrant"? (Johnson 2011)

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

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

(141) 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." (Maier 2015)

(142) 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." [*Cyprus Mail*, 30/5/11]

But we can go even further than that to account for data that are completely out of reach for other grammars. As we saw earlier (Sect. 7.3.1), given its psycholinguistically inspired nature, the DS-TTR model records the various alternative options arising during processing including those arising from the processing of ambiguous strings. Even options less probabilistically favoured, and hence not currently pursued, are stored temporarily in the context model (context DAG) in order to be employed for, e.g., the functioning of repair processes, like corrections, in dialogue (see e.g. Hough 2015; Eshghi et al. 2015). This independently needed modelling allows us to capture the variable semantic-"constituency" ambiguity of some mixed quotation strings and the ways they can be exploited by interlocutors, for example in puns and jokes (as pointed out by Maier 2014a):

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

Due to the fine-grainedness of the individual DS-TTR mechanisms and the non-differentiation of grammatical and pragmatic modes of processing, all the "peculiarities" of mixed quotation presented in Maier (2014a) and others (e.g. see earlier (138)–(139)) are eliminated here. This is because, in DS-TTR, there is no need to license a level of syntactic constituency or any independent syntactic categories for strings (see Gregoromichelaki in prep. for full formal implementation of particular instances).[37]

8 Conclusion

The view of NLs as codes mediating a mapping between "expressions" and the world has been abandoned here to give way to a view where utterances are seen as goal-directed actions aimed at locally and incrementally altering the affordances of the context for both one's self and one's interlocutors.[38] As conceived in the model presented here, such actions employ perceptual stimuli composed not only of words and syntax but also of elements like visual marks and styles, prosody, intonation and timing, gestures, facial expressions and gaze. All these aspects of the stimuli serve as triggers for the invocation not only of conceptual contents but also

[37] As an anonymous reviewer points out, there is potential for overgeneration in this overall approach. However, in my view, this should be handled on a case-by-case basis, given observed particularities of languages and constructions, not as systemic architectural grammatical constraints (unless there is solid evidence for the latter), which is what I am concerned with here.

[38] Goal-directedness should not be construed as consciously or even subconsciously "intentional" in the Gricean sense. All (subpersonal) DS-TTR grammatical operations are goal-directed in the sense that predictions of the next perceptual input are system-generated and, accordingly, constrain which input will be sought and how such input will be accommodated. For arguments against the Gricean construal see Gregoromichelaki et al. (2011), Gregoromichelaki (2013b), Gregoromichelaki et al. (2013b), and Pickering and Garrod (2004); see also Saka (2003/2005) for similar views regarding the processing of quotation.

time/space/psychological perspectives, remembered experiences, feelings, attitudes, beliefs, imagistic impressions etc. all of which constitute part of their "meaning". Thus, as part of this set, linguistic elements are not conceived as symbols and operations arbitrarily related to their referents and semantics. Instead they are seen as intrinsically linked to their phonetic or graphical realisations and the "meanings" they activate through human categorisation processes. From this perspective, any aspect of such stimuli can participate in the processes that constitute the "grammar", whose function is nothing else but the dynamic categorisation of various perceptual inputs and their integration with memory and action schemata in the process of generating the next action steps. This perspective does not allow for any process, like the alleged operation of "quotation", that segregates meaning from form, "demonstration" from reference, or syntax/semantics from pragmatics. During human interaction, due to the interlocutors' (partially) shared experiences and goals, perceptual inputs are able to trigger common action schemata, event invocations, and associations thus becoming the basis of joint performance coordination via the intersubjective affordances that they make available. From this point of view, linguistic knowledge is part of the abilities to coordinate effective interaction with the environment, one's own self, or one's interlocutors. In particular contexts, some of the various affordances that linguistic stimuli give access to will be more relevant than others in order to locally coordinate effective responses. Reporting, echoing, citing or metacommenting on aspects of the process itself are means through which some of the various aspects of meaningfulness can be foregrounded in the service of facilitating joint performance. It is not curious then that quotation bears common features with conversational phenomena: under the present view this is because it employs the same mechanisms as conversation, and consequently quotation is expected to interact with such conversational phenomena, e.g. repair and shared utterances, which also facilitate coordination. DS-TTR, in taking a psycholinguistically realistic action-grounded view of grammar, aims to model these interactions by subsuming quotation phenomena in a unified framework under general conversational coordinative mechanisms.[39]

References

Abbott, B. (2005). *Some notes on quotation*. In De Brabanter (2005b, 13–26).

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

[39]I wish to thank all my collaborators to the DS-TTR project: Ruth Kempson, Ronnie Cann, Stelios Chatzikyriakidis, Arash Eshghi, Pat Healey, Julian Hough, Chris Howes, Greg Mills, Matt Purver, and Graham White. I am especially grateful to Paul Saka for various suggestions, comments, invaluable editorial assistance and tremendous support. In addition, I am very thankful to an anonymous reviewer for insightful comments, data, and further sources of information. I acknowledge support from the ESRC (Grant ESRC-RES-062-23-0962).

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., Diaz, F., & Collins, A. F. (1996). Keeping your footing: Conversational completion in three-part sequences. *Journal of Pragmatics, 25*(2), 151–171.
Bateson, G. (1972). *Steps to an ecology of mind: Collected essays in anthropology, psychiatry, evolution, and epistemology*. Chicago: University of Chicago Press. 1987 \mbox{ISBN 0-87668-950-0.}
Bakhtin, M. M. (1981). The dialogic imagination: Four essays. In M. Holquist (Ed.). (C. Emerson & M. Holquist, Trans.). Austin: University of Texas Press.
Banfield, A. (1973). Narrative style and the grammar of direct and indirect speech. *Foundations of Language, 10*(1), 1–39.
Barwise, J., & Perry, J. (1983). *Situations and attitudes*. Cambridge, MA: MIT Press/Bradford.
Bonami, O., & Godard, D. (2008). On the syntax of direct quotation in French. In S. Mueller (Ed.), *Proceedings of the HPSG08 conference* (pp. 358–377). Stanford: CSLI Publications.
Brendel, E., Meibauer, J., & Steinbach, M. (2011). Introduction. In E. Brendel, J. Meibauer, & M. Steinbach (Eds.), *Understanding quotation*. Berlin: Mouton de Gruyter.
Brandom, R. (1994). *Making it explicit: Reasoning, representing, and discursive commitment*. Cambridge, MA: Harvard University Press.
Cann, R., Kempson, R., & Marten, L. (2005). *The dynamics of language*. Oxford: Elsevier.
Cann, R., Kempson, R., & Purver, M. (2007). Context and well-formedness: The dynamics of ellipsis. *Research on Language and Computation, 5*(3), 333–358.
Cappelen, H., & Lepore, E. (1997). Varieties of quotation. *Mind, 106*, 429–450.
Cappelen, H., & Lepore, E. (2007). *Language turned on itself. The semantics and pragmatics of metalinguistic discourse*. Oxford: Oxford University Press.
Chater, N., Pickering, M., & Milward, D. (1995). What is incremental interpretation? *Incremental interpretation, Edinburgh Working Papers in Cognitive Science*, 11.
Clark, H. H. (1996). *Using language*. Cambridge: 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.), *Handbook of the philosophy of science, vol 14: Philosophy of linguistics* (pp. 271–323). Amsterdam: Elsevier.
Cooper, R., & Ranta, A. (2008). Natural languages as collections of resources. In R. Cooper & R. Kempson (Eds.), *Language in flux* (pp. 109–120). London: College Publications.
Crystal, D. (2013). http://www.davidcrystal.community.librios.com/?id=2914. Accessed 13/1/15.
Cumming, S. (2005). *Two accounts of indexicals in mixed quotation*. In De Brabanter (2005b, 77–88).
Davidson, D. (1979). Quotation. *Theory and Decision, 11*, 27–40.
Davidson, D. (1984). *Inquiries into truth and interpretation*. Oxford: Clarendon Press.
De Brabanter, P. (2005a). *Introduction*. In De Brabanter (2005b, 1–2).
De Brabanter, P. (Ed.). (2005b). *Hybrid quotations*. Amsterdam: John Benjamins.
De Brabanter, P. (2010a). The semantics and pragmatics of hybrid quotations. *Language and Linguistics Compass*, *4*(2). https://doi.org/10.1111/j.1749-818X.2009.00185.x.
De Brabanter, P. (2010b). Uttering sentences made up of words and gestures. In B. Soria & E. Romero (Eds.), *Explicit communication* (pp. 199–216). London: Palgrave Macmillan.
De Brabanter, P. (2013). Francois Recanati's radical pragmatic theory of quotation. *Teorema, 32*, 109–128.
De Ruiter, J., Mitterer, H., & Enfield, N. J. (2006). Projecting the end of a speakers turn: A cognitive cornerstone of conversation. *Language, 82*(3), 515–535.
Duff, M. C., Hengst, J. A., Tranel, D., & Cohen, N. J. (2007). Talking across time: using reported speech as a communicative resource in amnesia. *Aphasiology, 21*(6–8), 702–716.

Eckardt, R. (2014). *The semantics of free indirect discourse: How texts allow us to mind-read and eavesdrop*. Leiden: Brill.

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)*, York, UK.

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

Eshghi, A., Hough, J., Purver, M., Kempson, R., & Gregoromichelaki, E. (2012). Conversational interactions: Capturing dialogue dynamics. In L. Borin & S. Larsson (Eds.), *From quantification to conversation*. London: College Publications.

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

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

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*. In De Brabanter (2005b, 109–28).

Gibson, J. J. (1977). The theory of affordances. In R. E. Shaw & J. Bransford (Eds.), *Perceiving, acting, and knowing*. Hillsdale: Lawrence Erlbaum Associates.

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*(3), 287–311.

Ginzburg, J., Fernández, R., & Schlangen, D. (2014). Disfluencies as intra-utterance dialogue moves. *Semantics and Pragmatics, 7*(9), 1–64.

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

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

Goodwin, C. (2007). Interactive footing. In E. Holt & R. Clift (Eds.), *Reporting talk: Reported speech in interaction* (pp. 16–46). Cambridge: Cambridge University Press.

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. (2012). Review of J. Ginzburg (2012) The interactive stance. *Folia Linguistica*, *47*(1), 93–316.

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*. Cham: Springer.

Gregoromichelaki, E., & Kempson, R. (2015). Joint utterances and the (split-) turn-taking puzzle. In *Interdisciplinary studies in pragmatics, culture and society* (pp. 703–743). Springer.

Gregoromichelaki, E., & Kempson, R. (2016). Reporting, dialogue, and the role of grammar. In *Indirect reports and pragmatics* (pp. 115–150). Springer.

Gregoromichelaki, E., & Kempson, R. (2017). Procedural syntax. In R. Carston, B. Clark, & K. Scott (Eds.), tba Cambridge University Press.

Gregoromichelaki, E., Sato, Y., Kempson, R. Gargett, A., & Howes, C. (2009). Dialogue modelling and the remit of core grammar. In *Proceedings of IWCS, Eighth international conference on computational semantics*.

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., & Cann, R. (2012). Language as tools for interaction: grammar and the dynamics of ellipsis resolution. *The Linguistic Review, 29*(4), 563–584.

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

Gregoromichelaki, E., Kempson, R., Howes, C., & Eshghi, A. (2013b). On making syntax dynamic: The challenge of compound utterances and the architecture of the grammar. In I. Wachsmuth, J. P. de Ruiter, P. Jaecks, & S. Kopp (Eds.), *Alignment in communication: Towards a new theory of communication*. Amsterdam: John Benjamins.

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. 213–235). Amsterdam: John Benjamins.

Horn, L. (1989). *A natural history of negation*. Chicago: University of Chicago Press.

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.

Johnson, M. (2011). The punctuation theory of quotation. In E. Brendel, J. Meibauer, & M. Steinbach (Eds.), *Understanding quotation*. Berlin: Mouton de Gruyter.

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: Mathematical Centre Tracts.

Kamp, H., & Reyle, U. (1993). *From discourse to logic*. Dordrecht: Kluwer.

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

Kempson, R., Gregoromichelaki, E., Purver, M., Mills, G. J., Gargett, A., & Howes, C. (2009a). How mechanistic can accounts of interaction be? In *Proceedings of Diaholmia, the 13th workshop on the semantics and pragmatics of dialogue*.

Kempson, R., Gregoromichelaki, E., & Sato, Y. (2009b). Incrementality, speaker/hearer switching and the disambiguation challenge. In *Proceedings of European Association of Computational Linguistics* (EACL). Athens.

Kempson, R., Gregoromichelaki, E., & Howes, C. (2011a). *The dynamics of lexical interfaces*. Chicago: CSLI Publications.

Kempson, R., Gregoromichelaki, E., Meyer-Viol, W., Purver, M., White, G., & Cann, R. (2011b). 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, April 26–27). Joint utterances in Greek: Their implications for linguistic modelling. In *Proceedings of 33rd annual linguistics meeting "Syntactic Theories and the Syntax of Greek"*. Thessaloniki

Kempson, R., Cann, R., Eshghi, A., Gregoromichelaki, E., & Purver, M.. (2015). Ellipsis. In S. Lappin & C. Fox (Eds.), *Handbook of contemporary semantics* (2nd ed., pp. 114–140).

Kempson, R., Cann, R., Gregoromichelaki, E., & Chatzikyriakidis, S. (2016). Language as mechanisms for interaction. *Theoretical Linguistics, 42*(3–4), 203–276. https://doi.org/10.1515/tl-2016-0011.

Kempson, R., Cann, R., Gregoromichelaki, E., & Chatzikyriakidis, S. (2017). Action-based grammar. *Theoretical Linguistics*, 43(1–2), 141–167.

Koder, F., & E. Maier (2014). Children mix direct and indirect speech: Evidence from pronoun comprehension. Ms. University of Groningen.

Larsson, S. (2011, September). The TTR perceptron: Dynamic perceptual meanings and semantic coordination. In *Proceedings of the 15th workshop on the semantics and pragmatics of dialogue* (pp. 140–148). SemDial 2011 – Los Angelogue.

Larsson, S. (2015). Formal semantics for perceptual classification. *Journal of Logic and Computation, 25*(2), 335–369.

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.

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

Lerner, G. H. (1996). On the "semi-permeable" character of grammatical units in conversation: conditional entry into the turn space of another speaker. In E. Ochs, E. A. Schegloff, & S. Thompson (Eds.), *Interaction and grammar* (pp. 238–276). Cambridge: Cambridge University Press.

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

Maier, E. (2014a). Mixed quotation: The grammar of apparently transparent opacity. *Semantics & Pragmatics, 7*(7), 1–67.

Maier, E. (2014b). Language shifts in free indirect discourse. *Journal of Literary Semantics, 43*(2), 143–167.

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

Maier, E. (2017). Mixed quotation. Ms. University of Groningen (survey article for The Companion to Semantics edited by Lisa Matthewson, Cécile Meier, Hotze Rullmann, Thomas Ede Zimmermann, Wiley/Blackwell, New York.)

Marslen-Wilson, W. (1973). Linguistic structure and speech shadowing at very short latencies. *Nature (London), 244*, 522–523.

McCawley, J. D. (1999). Participant roles, frames, and speech acts. *Linguistics and Philosophy, 22*(6), 595–619.

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, DC: Georgetown University Press.

Morris, P. (1997). *The Bakhtin reader: Selected writings of Bakhtin, Medvedev, Voloshinov*. Oxford: Oxford University Press.

Noh, E.-J. (1998). Echo questions: Metarepresentation and pragmatic enrichment. *Linguistics and Philosophy, 21*, 603–628.

Norrick, N. (2015). Indirect reports, quotation, and narrative. In A. Capone, F. Kiefer, & F. Lo Piparo (Eds.), *Indirect reports and pragmatics*. Dordrecht: Springer.

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

Pickering, M., & Garrod, S. (2004). Toward a mechanistic psychology of dialogue. *Behavioral and Brain Sciences, 27*(02), 169–190.

Pickering, M., & Garrod, S. (2013). An integrated theory of language production and comprehension. *Behavioral and Brain Sciences, 36*(4), 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), 309–347.

Poesio, M., & Traum, D. (1998). Towards an axiomatization of dialogue acts. In *Proceedings of the twente workshop on the formal semantics and pragmatics of dialogues* (13th twente workshop on language technology).
Postal, P. M. (2004). *Sceptical linguistic essays*. Oxford University Press.
Potts, C. (2007). The dimensions of quotation. In P. Jacobson & C. Barker (Eds.), *Direct compositionality* (pp. 405–431). Oxford: Oxford University Press.
Predelli, S. (2003). Scare quotes and their relation to other semantic issues. *Linguistics and Philosophy, 26*, 1–28.
Purver, M., Cann, R., & Kempson, R. (2006). Grammars as parsers: Meeting the dialogue challenge. *Research Language Computation, 4*, 289. https://doi.org/10.1007/s11168-006-9007-x
Purver, M., Howes, C., Gregoromichelaki, E., & Healey, P. G. T. (2009). Split utterances in dialogue: A corpus study. In *Proceedings of SigDial*. London.
Purver, M., Gregoromichelaki, E., Meyer-Viol, W., & Cann, R.. (2010). Splitting the *I*'s and crossing the *you*'s: Context, speech acts and grammar. In *Proceedings of SemDial 2010* (PozDial), Poznan, Poland.
Purver, M., Eshghi, A., & Hough, J. (2011). Incremental semantic construction in a dialogue system. In *Proceedings of the 9th international conference on computational semantics* (IWCS). Oxford.
Quine, W. V. O. (1940). *Mathematical logic*. Cambridge, MA: Harvard University Press.
Ranta, A. (1994). *Type-theoretical grammar*. Oxford: Clarendon Press.
Recanati, F. (2000). *Oratio Recta, Oratio Obliqua*. Cambridge, MA: 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.
Ross, J. R. (1970). Metalinguistic anaphora. *Linguistic Inquiry*, 1, 273.
Kempson, R., Gregoromichelaki, E., Eshghi, A., & Hough, J.. (to appear) Ellipsis in Dynamic Syntax. In J. van Craenenbroeck & T. Temmerman, (Eds.), *The Oxford handbook of ellipsis*. ISBN: 9780198712398
Sacks, H., Schegloff, E. A., & Jefferson, G. (1974). A simplest systematics for the organization of turn-taking for conversation. *Language, 50*, 696–735.
Sag, I. A., Wasow, T., & Bender, E. M. (2003). *Syntactic theory*. Stanford: CSLI Publications.
Saka, P. (1998). Quotation and the use-mention distinction. *Mind, 107*(425), 113–135.
Saka, P. (2003/2005). Quotational constructions. *Belgian Journal of Linguistics*, 17(1), 187–212.
Saka, P. (2011). Quotation and conceptions of language. *Dialectica, 65*(2), 205–220.
Saka, P. (2013). Quotation. *Philosophy Compass, 8*(10), 935–949.
Sato, Y. (2011). *Local ambiguity, search strategies and parsing in dynamic syntax*. In Kempson et al. (2011b) (Eds.).
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.
Shan, C.-c. (2011). The character of quotation. *Linguistics and Philosophy, 33*(5), 417–443.
Sharvit, Y. (2008). The puzzle of free indirect discourse. *Linguistics and Philosophy, 31*(3), 353–395.
Slama-Cazacu, T. (1976). Nonverbal components in message sequence: 'mixed syntax'. In W. C. McCormack & S. A. Wurm (Eds.), *Language and man: Anthropological issues* (pp. 217–227). The Hague: Mouton.
Steedman, M. (1992). *Grammars and processors* (Tech. Rep. MS-CIS-92-52). University of Pennsylvania, Department of Computer and Information Science.
Stone, M. (2004). Intention, interpretation and the computational structure of language. *Cognitive Science, 28*(5), 781–809.
Tanenhaus, M. K., Spivey-Knowlton, M. J., Eberhard, K. M., & Sedivy, J. C. (1995). Integration of visual and linguistic information in spoken language comprehension. *Science, 268*, 1632–1634.

Tannen, D. (1986). Introducing constructed dialogue in Greek and American conversational and literary narratives. In F. Coulmas (Ed.), *Direct and indirect speech* (pp. 311–322). Berlin: Mouton.

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.