14 Procedural Syntax

Eleni Gregoromichelaki and Ruth Kempson

14.1 Introduction

Within psychological linguistic theories, relevance theory (RT, Sperber & Wilson 1986/1995), with its focus on explaining how communication works, early on directed attention to general cognitive processes, notably an efficiency-driven processor striving to balance cognitive goals against processing effort. In this chapter, we will argue that taking such a goal-directed processing approach can be extended much deeper so that fundamental natural language (NL) properties, namely, the structuring of NL signals, can be explicated. We are going to introduce a grammar framework, Dynamic Syntax, which, in conceptualising NLs as procedures for (inter)action, directly provides an alternative 'knowing-how' motivation of NL structure. The top-down predictive articulation of DS can be seen as a radical extension of the concept of *procedure* widely adopted in RT (Blakemore 1987, 2002; Wilson, 2000, 2011b). Under this extension, procedures are not some add-on to an otherwise standard form–meaning encoding but a total replacement of the code-based model. We believe that this approach, despite its novelty, is compatible with Wilson (2016), where the notion of 'procedure' is cashed out in non-representational dispositional terms following Bezuidenhout (2004). However, a challenge that then arises is how to account for the reification of such procedures as exemplified in apparent metarepresentational practices like quotation, reporting, citation etc. We will argue that even such phenomena can receive adequate and natural explanations through a grammar that allows for the dynamic ad hoc creation of occasion-specific content through self-referential mechanisms.

14.2 Split Utterances: The Challenge of Incrementality

A ubiquitous feature of informal conversations is how little exchanges rely on the formulation of the complete, well-formed sentences/thoughts that standard grammars take as their remit (see e.g. Miller & Weinert 1998). Instead, people opportunistically rely on the linguistic and non-linguistic context for what to say and understand, coordinating verbal and non-verbal actions, and

handling efficiently what have been considered elsewhere as 'degenerate' inputs (Radford 2004; Chomsky 2006):

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

Thus, linguistic behaviour in dialogue undermines basic theoretical notions like the abstract concepts of 'sentence' and 'syntactic constituency': interlocutors can take over from each other at any point in a clausal sequence across all syntactic and semantic dependencies, change or abandon strings while incomplete, and still manage to coordinate successfully. Stainton (2006) presents evidence that standard syntactic/semantic accounts cannot deal with various types of 'fragments' in conversation through ellipsis analyses and, instead, an RT-type pragmatic account is required:

(14.2) A: Covent Garden?
B: Right at the lights. Then straight on up.

However, the phenomenon is much more extended and problematic. Firstly, contra Stainton, in many languages, there are syntactic/morphological restrictions constraining the form of non-sentential strings used to perform various speech acts like clarifications/corrections/etc. (Ginzburg 2012; Gregoromichelaki 2012). Therefore, syntax/semantics must be involved in their licensing. Moreover, all such speech acts can be perfectly processable well before any sentential or propositional unit has been delivered by standard grammars or any informative intention-grounding pragmatic inference has been derived:

(14.3) A: Er, the doctor
B: Chorlton?
A: Chorlton, mhm, he examined me, erm, he, he said now they were on about a slight [shadow] on my heart.
[BNC: KPY 1005–1008]

For many other cases, just abandoning syntactic/semantic accounts and relegating the analysis to pragmatics doesn't help. This is because the phenomenon illustrated in (14.1) earlier and below, so-called split utterances (SUs), involves

syntactic/semantic constraints operating across shifts of contextual parameters, e.g. speaker/hearer roles. For example, (14.4) involves a split between a preposition and its complement and (14.5) a split between determiner and noun:[1]

(14.4) JACK: I just returned
KATHY: from…
JACK: Finland. [Lerner 2004]

(14.5) A: I need a a…
B: mattock. For breaking up clods of earth. [BNC]

Such split dependencies can be arbitrarily complex. (14.6) splits apart both a dependency between the left-peripheral *wh*-form and the associated object 'gap' position of *axe* as well as the auxiliary and verb; (14.7) separates the negative polarity *any of the saliva kits* from A's initiating utterance on which it is dependent across B's intervention, while simultaneously being construed as an extension of B's intervention:

(14.6) A: Which unit are we thinking we should…
B: axe? None.

(14.7) A: Has every female gymnast handed in
B: her blood sample?
A: or even any of the saliva kits?

But SU data are also problematic for all syntactic/semantic frameworks that rely on linear-order independent, bottom-up licensing of syntactic/semantic constituency and dependencies (see e.g. Hornstein et al. 2005; cf. Putnam 2009, 2010; Osborne et al. 2011). Given that such frameworks rely on a notion of 'sentence' or 'phrase' (see e.g. Collins 2003, 2007) as the minimal units that can be interpreted, they must necessarily resort to either an assumption of ellipsis regarding each substring or a supplementary competence/performance division of labour (Kobele 2016; cf. Kempson et al. 2017, 2019) in order to account for such data, if at all.

An ellipsis account is not viable since encoded deletion operations will have to be postulated for any partial string, for example, missing subjects in English:

(14.8) HESTER COLLYER: It's for me.
MRS ELTON THE LANDLADY: And Mr. Page?

[1] Such phenomena are also beyond the reach of analyses like Barton (1990) and Barton and Progovac (2005) since these accounts, despite their welcome pioneering contribution, concentrate and rely exclusively on predefined, well-formed non-sentential constituents and not the role of incremental licensing.

HESTER COLLYER: is not my husband. But I would rather you continue to think of me as Mrs. Page.
[from *The Deep Blue Sea* (film)]

Beyond postulating ellipsis, resorting to a distributed competence/performance account not only does not deliver a parsimonious explanation it also appears empirically inadequate. This is because the problems for syntactic/semantic frameworks are not just due to the bottom-up licensing imposed by standard competence approaches. It is the very notion of syntactic constituency, mediating between pairings of phonological–semantic representations (Brody 2002), that gets in the way. This remains the case even when non-standard constituents are employed (Steedman 2000) since the SU-splits can occur at any point. Any constituency constraints included in the competence account preclude an intuitive systematic analysis of SUs as simple continuations, because even invoking shared structures that somehow splice together surface syntactic structures will, in many cases, result in ill-formed strings or incorrect interpretations:

(14.9) A [SEEING B EMERGE FROM A SMOKE-FILLED KITCHEN]: Did **you** burn
B [INTERRUPTING]: **myself**? No, fortunately not.

Moreover, note that SU partial strings are incrementally interpretable since they can be used to accomplish speech acts (see (14.8)) *necessarily* without any propositional enrichment (see (14.10)), otherwise their incompleteness will not be adequate as a parsimonious explanation for the triggering of the completion:

(14.10) PSYCHOLOGIST: And you left your husband because …
CLIENT: we had nothing in common anymore

Even frameworks that integrate syntax/semantics and pragmatics within the grammar (Ginzburg 2012; Ginzburg et al. 2014) do not provide a unified account of all non-sentential cases since they separate out some 'fragments' to be treated as propositional, resolved through powerful coercion operations, and others as continuations either by the speaker ('self-repair') or the hearer ('completions', Poesio & Rieser 2010). The problem that permeates such accounts is the necessary encoding of higher-order speech-act characterisations (e.g. 'The speaker is asserting/querying/clarifying x') associated with the derivation of propositions for the licensing of each contextual switch in dialogue. However, this does not provide for the intuitive unified explanation of the phenomena, which, in our view, is the potential for top-down predictive, incremental, and interactive processing. The ability to incrementally set out

and resolve dependencies by either interlocutor, instead of appealing to some constituency-defining competence, underlies the successful coordination of speaker/hearer actions displayed in (14.1)–(14.10). And this ability underpins monologue as well as dialogue, as evidenced by phenomena like parenthetical insertion:

(14.11) Well, they dropped cards in *I suppose* the doors (Dehé 2014: 65)

(14.12) Hi, and welcome to New Books in Sociology, a podcast where we interview authors of interesting and influential books in the field of, *you guessed it*, sociology…

[Sociology-Halperin]

Immediate, sub-sentential coordination among interlocutors, instead of the communication of codified constituent structures,[2] explains why interlocutors can achieve what from an analyst's external point of view can be described as propositional 'speech acts' without the interlocutors needing to process either full sentences or propositions. Regarding the mechanisms available to the interlocutors, such higher-order explicatures (e.g. 'The speaker is asserting/claiming/querying/requesting *p*') are not a necessary precondition for coordination.[3] For example, in (14.8) and (14.10) earlier, initiating a dependency is all that is needed to induce a resolving continuation by the other interlocutor. Interlocutors do not, in addition, need to infer/plan following higher-order speech acts/intentions (Gregoromichelaki et al. 2011, 2013a; see also Wilson & Wharton 2006). Meaningful coordination in interactive exchanges needs to be locally opportunistic (Clark 1996) so that efficiency requires no essential guidance by some preformed overarching plan by either participant. This is more evident in non-cooperative cases where seamless and timely intervention is perfectly possible without any necessary consideration of the intentions of the first speaker:

(14.13) A: It's obvious from what he says (that)
B: (that) you are wrong.

[2] Codified constituent structures occur, for example, in dialogue modelling where the grammar includes derivations of 'I am asking you whether you intended to utter x' for clarification requests.

[3] This is also argued within RT (Sperber & Wilson 1986/1995: 244) for specific speech-act characterisations. However, we go further in that we claim that not even a 'proposition expressed' is a necessary ingredient of coordination/communication and weak communicative acts like 'saying' (Sperber & Wilson 1986/1995: 247) need to be analysed not as deriving propositional schemata but as descriptions of the fundamental (turn-taking) procedures induced by the grammar (Eshghi et al. 2015; Gregoromichelaki & Kempson 2015; Gregoromichelaki 2017).

SU exchanges thus demonstrate the exercise of very basic human abilities, 'knowing how to go on' (Wittgenstein 1953: Section 150) in any process, so that even very young children participate – indeed SUs form the staple diet upon which they learn language, as displayed by the Old MacDonald nursery rhyme (see also Arnon et al. 2014):

(14.14) A: Old MacDonald had a farm, E-I-E-I-O and on that farm he had a
B: cow
A: and the cow goes
B: moo

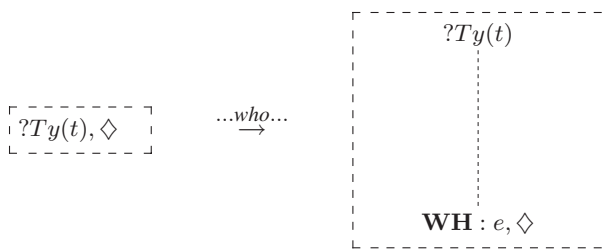
In conclusion, the seamless fluency with which interlocutors take on or hand over utterance responsibility presents a formidable challenge to current formalisms where syntax/semantics are defined as operating bottom-up with reference to constituents, with these then feeding into some extra-grammatical account of pragmatic inference. For these reasons, we turn now to a view of the grammar that essentially employs a view of syntax/semantics as goal-directed predictive action (procedures) for interactive coordination, not exchange of propositions, with no separation imposed between some putative grammatical, as opposed to extra-grammatical, context and structure.

14.3 Dynamic Syntax

Dynamic Syntax (Kempson et al. 2001; Cann et al. 2005) is a grammar architecture whose core notion is incremental interpretation of word sequences (comprehension) or linearisation of contents (production) relative to context. The Dynamic Syntactic engine, including the lexicon, is articulated in terms of goal-driven actions that are accomplished by giving rise to expectations of further actions, by processing contextual input, or by being abandoned as unviable in view of more competitive alternatives. Thus words, syntax and morphology are all modelled as 'affordances', opportunities for action that interlocutors can deploy to perform step-by-step a mapping from perceivable stimuli (phonological strings) to conceptual structure, or vice versa.

To illustrate, we display in (14.15)–(14.18) the (condensed) steps involved in the parsing of a long-distance dependency:

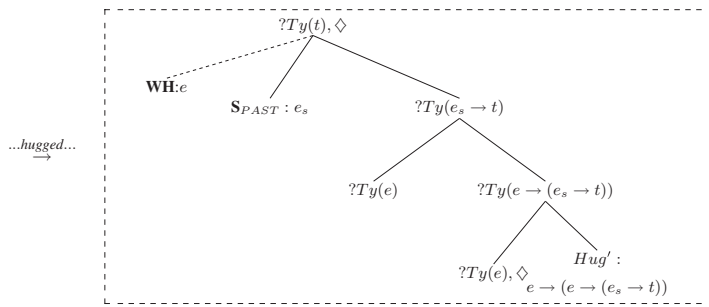
(14.15)



The task always starts with a set of probabilistically weighted predicted information states (we show only one and only the syntactically relevant part). Such states include routinised goals to build conceptual structures of some ontological type (e for entities in general, e_s for events, $e \rightarrow (e_s \rightarrow t)$ for predicates, etc.). In (14.15), the goal to build a proposition of type t is shown as a one-node tree with the goal *?Ty(t)* and the current focus of attention, the pointer ◊. The pointer at a node including a goal drives the prediction of further affordances/subgoals, expected to eventually satisfy the current goal either by the processing of (verbal) input (as a hearer) or by producing that input (as a speaker). For (14.15), one of the probabilistically licensed next steps for English (executed by lexical and general computational *macros* of actions) is illustrated next: a prediction that a structurally underspecified node (indicated by the dotted line) can be built and can accommodate the result of parsing/generating *who*. As illustrated here, temporary uncertainty about the eventual contribution of some element is implemented through *structural underspecification*. Initially unfixed tree-nodes model the retention of the contribution of the *wh*-element in a memory buffer until it can unify with some argument node in the upcoming local domain. Non-referential words like *who* and other semantically underspecified elements (e.g. pronominals, anaphors, auxiliaries, tenses) contribute *underspecified content* in the form of so-called *metavariables* (indicated in bold font), which trigger search for their eventual type-compatible substitution from among contextually salient entities or predicates.

General computational and lexically triggered macros then intersperse to develop a tree: in (14.16), the verb contributes both conceptual structure in the form of unfolding the tree further and fetching an ad hoc concept (indicated as *Hug'*) developed according to contextual restrictions,[4] as well as placeholder metavariables for time and event entities to be supplied by the context. Finally, the pointer ◊ is left at the argument node implementing the word-order restriction that the object needs to follow the verb in English:

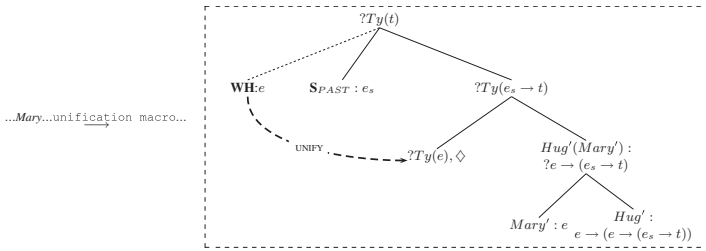
(14.16)



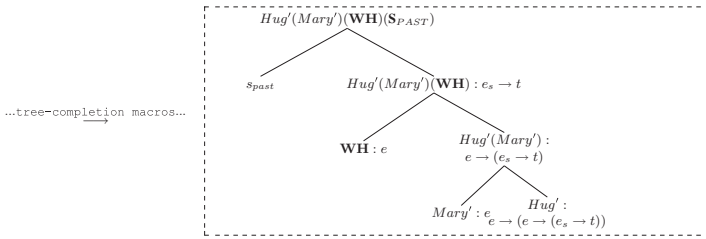
[4] For reasons of space, we present here a very simplistic view of conceptual content as atomic formulae; for more extensive views, see Wilson and Carston (2007); Carston (2012, 2013, 2016a); Cann and Kempson (2017); Gregoromichelaki (2017).

At this point, the word *Mary* can be processed to deliver a contextually determined individual (*Mary'*) at the argument node of the predicate. After this step, everything is in place for the structural underspecification to be resolved, namely, the node annotated by *who* can now unify with the subject node of the predicate, which results in a well-formed explicature representing the minimal content of an utterance of *Who hugged Mary?* in binary tree form:[5]

(14.17)



(14.18)



In this instance, processing a single speaker's utterance has ended with unresolved content, modelled as a metavariable needing substitution from context, the contribution of the *wh*-pronoun. The substitution of such a metavariable can be provided either by the same speaker (e.g. as in a rhetorical question, self-questioning/correction, etc.) or by the interlocutor. Besides trees, the complete model includes the full set of licensed next steps for partial trees currently entertained as live options, a record of activated entities/events/propositions, and contextual parameters like the utterance/locutionary event, speaker-hearer roles, and times. All these elements constitute an *information state* (Hough 2014; Gregoromichelaki 2017). The information state tracks not only the shifting parameters of each word-utterance event serving for the interpretation of various indexical metavariables but also the actions used, recorded as traversals of paths in a graph display, and the processing paths that have been considered as probabilistically live

[5] For a fuller discussion of the framework, see Kempson et al. (2001); Cann et al. (2005); Eshghi et al. (2010, 2011, 2012, 2015); Kempson et al. (2016, 2017).

options but not eventually pursued (see e.g. Sato 2011; Hough 2014). Storing the action paths is necessary for the resolution of anaphora and ellipsis, especially 'sloppy' readings:

(14.19) A: John upset his mother.
B: Harry too.

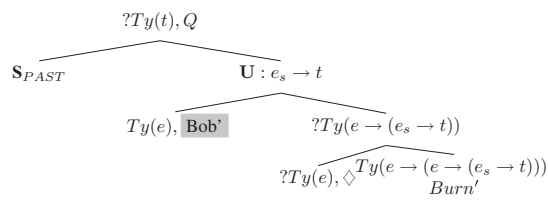
Maintaining abandoned options is required for the explicit modelling of conversational phenomena like clarification, self-/other-corrections, etc. but also for humorous effects and puns (Gregoromichelaki 2017).

The DS model assumes tight interlinking of NL perception and action. The predictions generating the sequence of trees above are equally deployed in comprehension and production. Comprehension involves the generation of predictions/goals and awaiting input to satisfy them, while production involves the deployment of action (verbalising) by the predictor themselves in order to satisfy their predicted goals. By imposing top-down predictive and goal-directed processing at all comprehension/production stages, interlocutor feedback is constantly anticipated and seamlessly integrated (Gargett et al. 2008, 2009; Purver et al. 2010; Eshghi et al. 2015; Gregoromichelaki 2017). At any point, either interlocutor can take over to realise such predicted goals. This can be illustrated in the sharing of the dependency constrained by the locality definitive of reflexive anaphors:

(14.20) MARY: Did **you** burn
BOB: **myself**? No.

As shown in (14.20), a simplified version of (14.9), Mary starts a query involving an indexical metavariable contributed by *you* that is resolved by reference to the *Hearer'* contextual parameter currently occupied by *Bob'*:

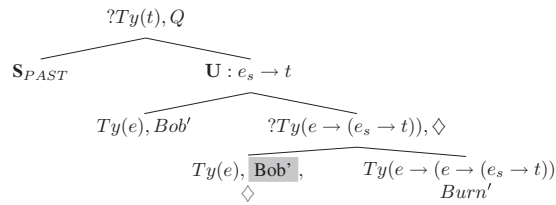
(14.21) *Mary:Did you burn* ⟼



With the information state tracking the speaker/hearer roles as they shift sub-sententially, these roles are reset in the next step when Bob takes over the utterance. *Myself* is then uttered. Being a pronominal, it contributes a metavariable and, being a reflexive indexical, it imposes the restriction that the entity to substitute that metavariable needs to be a co-argument that bears the

Speaker' role. At this point in time, the only such available entity in context is again *Bob'*, which is duly selected as the substituent of the metavariable:

(14.22) *Bob*: *myself* ?



As a result, binding of the reflexive is semantically appropriate, and locality is respected even though joining the string as a single sentence would be ungrammatical. This successful result relies on the lack of a syntactic level of representation and the sub-sentential licensing of contextual dependencies which renders the fact that the utterance constitutes a joint action irrelevant for the wellformedness of the string. Concomitantly, coordination among interlocutors here can be seen, not as propositional inferential activity, but as the outcome of the fact that the grammar consists of a set of licensed complementary actions that speakers/hearers perform in synchrony (Gregoromichelaki et al. 2011, 2012; Gregoromichelaki 2013b; Gregoromichelaki & Kempson 2015). Due to sub-sentential step-by-step licensing, speakers are not required to plan propositional units, so hearers do not need to reason about propositional intentions. Given that parsing/production are predictive activities, a current goal in the information state may be satisfied by a current hearer, so that it yields the retrieval/provision of conceptual information that matches satisfactorily the original speaker's goals, as in (14.5)–(14.7), (14.9), or is judged to require some adjustment that can be seamlessly and immediately provided:

(14.23) KEN: He said 'the married couple will walk- walk down the street and they will be all dressed up and people will come by with.hh
LOUISE: rice.
KEN: rice, petals or anything they think is suitable'. [Lerner (2004) modified]

However, especially in confrontational situations as in (14.13), there is no need to assume that the hearer expends any effort to recognise the speaker's intention since completion by the interrupting party is a means of avoiding having to achieve any such recognition. So, what the shift of theoretical perspective has achieved in making procedures central to the processing mechanism qua grammar is a shift in the burden of explanation. No longer do all communicative

effects need to be explained inferentially by grammar-external higher-order inference operating on representations of speaker intentions; instead, the interactive 'mind-reading' effect emerges from production/parsing involving 'predictive processing',[6] defined within the grammar itself. However, we can go even further than this, to cases of alleged explicit metarepresentation, and see how such effects can be achieved within a grammar appropriately conceived.

14.4 Quotation: Metarepresentation or Demonstration?

As we saw earlier in (14.4), (14.8) and (14.10), perfectly intelligible moves in dialogue can be achieved simply by initiating a grammatical dependency which prompts either interlocutor to fulfil it without specific determination or identifiability of a given speech act/intention. In various other cases, though, the interlocutor completing somebody else's utterance might be perceived as offering the completion along with a query as to whether such a (meta) representation is what the other interlocutor would have said (e.g. (14.6)–(14.7)). There are further so-called 'metarepresentational'/echoing phenomena in cases of citation, quotation, reports, echoic uses/irony, and code-switching:

(14.24) 'Cities,' he said, 'are a very high priority.' [FrameNet]

(14.25) **Wright** won't disclose how much the Nike deal is worth, saying only that 'they treat **me** well'. (The Face, 9–93:55) [De Brabanter 2010]

(14.26) A doctor tells him [Gustave Flaubert] he is like a 'vieille femme hysterique' (TLS online: 18-12-1998) [De Brabanter 2010]

(14.27) Alice said that life is 'difficult to understand'. [Cappelen & Lepore 1997]

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

Despite recent attempts to integrate such phenomena within standard grammars (e.g. Potts 2007b; Ginzburg & Cooper 2014; Maier 2014a,b), certain data are not amenable to appropriate treatment due to incrementality not being included as an explanatory factor within these formalisms. For example, as can be seen above, quotation can appear sub-sententially, and discontinuously, at any point, which means that contextual parameters characterising the utterance as echoic (Noh 1998, 2000; Wilson 2000) need to be able to shift incrementally at each word-by-word processing stage. Split utterances are one of the environments where quotation occurs frequently (Lerner 1991)

[6] Prediction is argued to be an essential part of human cognition in general (e.g. Friston 2010; Clark 2016).

due to the opportunity for co-constructing a joint perspective on some (actual or imaginary) speech/thought eventuality:

(14.29) 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]

To analyse such cases, the grammar needs to include the resources that take into account embedded and sub-sentential context switches since such indexicals do not track the actual current speaker/hearer roles. Moreover, such sub-sentential switches include cases where the same structure can be employed both as expressing a speaker's own voice and as a subsequent quotation:

(14.30) 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]

(14.31) 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 constructions and speech-act responsibility and echoing (Goodwin 2007; Kempson et al. 2011; Gregoromichelaki 2013a; Gregoromichelaki & Kempson 2016). For these reasons, a unified account of the synergy of such NL devices to derive appropriate contents cannot be given within standard theories involving encapsulation and division of labour between semantics/pragmatics and syntax.

Gregoromichelaki (2017) argues that such global interactions are straightforwardly accounted for in DS due to its incremental modelling of context shifting, the potential for sharing of syntactic/semantic dependencies, and the fact that there is no requirement to derive a comprehensive propositional speech act. However, modelling the potential of partially assuming another speaker's role, being perceived as 'demonstrating' what somebody was going to say/think, and the 'metalinguistic' appearance of phenomena like the quoted strings in (14.32)–(14.34) might seem especially problematic for the DS approach that eschews syntactic representations at the level of strings:

(14.32) 'La neige est blanc' is well-formed/grammatical/correct.

(14.33) James says that 'Quine' wants to speak to us. [James thinks that McPherson is Quine, Recanati 2010]

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

Unlike RT, as we saw, DS takes words/constructions as *affordances*, triggers for action, exploited by speakers/hearers to achieve coordination. This means that words/constructions are not defined as abstract code elements, expression types that are associated with referential/semantic values (representations).[7] With no expression types to serve as the semantic values of quotations (as metarepresentations), only mechanisms for processing stimuli, quotation thus offers a crucial test for the legitimacy of DS claims: when processing a quoted/cited string, what happens within the quotation marks (or any other indications), if not metarepresentation (Wilson 2000), following DS assumptions?

Gregoromichelaki (2017) argues that these cases are also unproblematic for DS and need not be handled through metarepresentation as in RT,[8] even though (some of them) can be treated as echoic. This account employs the extended DS-TTR model (Purver et al. 2010; Hough 2014), which models conceptual structure as tree-structures annotated with so-called *record-types* (Type-Theory-with-Records, TTR, (Cooper 2012)), that is, complex type values modelling the fine-grained structure of concepts. The advantage of TTR is that types (concepts) and dependencies among them can be constructed ad hoc instead of being pre-specified in the model (following also Carston 2002; Cann & Kempson 2017). Through the subtyping mechanism, conceptual underspecification and progressive enrichment can be implemented formally. In combination with the incremental DS syntactic engine, DS-TTR then takes a *grammar* as just the time-linear procedures of conceptualising/manifesting, elaborating, and reasoning over contextually integrated stimuli (not necessarily linguistic ones). In turn, this now provides the potential to explicate the broader family of echoic/metalinguistic phenomena in a natural manner that conforms with intuitions and parallels the DS modelling of anaphora/ellipsis.

First, in order to model cases like (14.6)–(14.7) and (14.24)–(14.29), as well as mid-sentence code-switching in general, it is assumed that the predictions driving processing at each step are induced by variable categorisation/linearisation systems (*grammars*/idiolects) that can shift at each word-utterance stage. Such shifts are indicated by switching the value of the designated metavariable in the information state that keeps track of which and whose grammar is being employed at each sub-sentential stage (see also Recanati 2010: 256; Ginzburg & Cooper 2014). Next, consider the most challenging cases, namely, metalinguistic uses, for example (14.32), where an NL-string appears in a

[7] For a notion of 'representation' approximately compatible with the DS claims, see e.g. Egan (2014); Bickhard (2016); Wiese (2017). According to DS, representational abstractions of the actual mechanics of conversation may be involved in explicit reasoning that can ensue in cases of failure of the fundamentally subpersonal cognitive mechanisms, deception, manipulation, etc.

[8] Contra Wilson (2000), Gregoromichelaki (2017) argues that all such cases can be analysed as 'demonstrations', even metalinguistic cases, as long as an appropriate code-free conception of the grammar is adopted.

regular NP/DP position. Under DS assumptions, this will be a pointer position where the English grammar ($DS_{English}$) has already generated a prediction for the processing of a singular term, an entity type (*?Ty(e)*). However, the grammar does not specify how such an entity type is to be provided and there are always multiple ways to derive such contents. The explanation of how such content is derived here is based on the independent assumption that actions/procedures are first-class citizens in DS. This means that the subpersonal grammatical mechanisms include (also subpersonal) higher-order operations aggregating and (re-)running chunks of actions. This is needed, for example, in order to provide parallel but distinct contents in the anaphoric cases of 'sloppy readings' (see earlier (14.19)), deep anaphora, etc. The idea now is that the same mechanism can be used cataphorically for metalinguistic, echoic, and similar uses combined with two additional DS-TTR assumptions: (a) a grammar maps stimuli (eventualities) to conceptual types or vice versa, and (b) new types (concepts) can be constructed and manipulated ad hoc on-line by enriching the available TTR annotations.

Under these assumptions, (14.34) illustrates a case where the sequence of actions a speaker performs under the guidance of *some* grammar (here French) is isolated from the rest of the processing task (implemented by a change in contextual parameters) and conceptualised/'reified' as a subtype of the type 'entity' (*Ty(e)*), an eventuality entity ($Ty(e_s)$). This eventuality can then satisfy the predicted goal, (*?Ty(e)*), to be achieved on the node where the pointer resides following $DS_{English}$. By assumption, a grammar maps eventualities to conceptual categorisations, but in the usual case only the conceptual structure is retained long term while the actual action execution is stored separately and decays fast. For anaphora/ellipsis resolution, the grammar also includes instructions to build content from any type of stimulus. In (14.34), the speaker's performance, the speaker's actual execution of a sequence of DS_{French} actions (literally a 'demonstration' of a speech event), is performed while a $DS_{English}$ prediction for a singular term is in force. Thus, the performance itself becomes part of the conceptual structure derived with the predicted $DS_{English}$-type generated on-line as the performance unfolds. The rest of the string then, with the speaker having moved out of the DS_{French}-demonstration mode, delivers a content that characterises in some way the DS_{French}-demonstration event. This delivers the interpretation that the event of using the French grammar in this way can be characterised as in accordance with the speaker's grammaticality/correctness judgements.[9]

For the cases where the interpretation of the indexicals reflects an echoed utterance/thought event, e.g. (14.25), (14.27)–(14.29), (14.33)–(14.34),

[9] Note that DS_{French} is *not* part of the speaker's awareness. The grammaticality/correctness judgement refers to some approximation of the actual DS_{French}.

similarly, the relevant nodes where the strings in quotation marks are processed are annotated through demonstrations the speaker performs. The only addition to the previous process is that now the invoked action execution, besides a change in the grammar, also involves a shift in the values of contextual parameters which are now supplied by some salient utterance/thought event.

Cases of *direct quotation* where a verb, e.g. *say*, requires a complement presented as an echoed utterance/thought eventuality are analysed as cases where such a contextual switch has been grammaticalised. Such verbs initially introduce a metavariable as a placeholder for their complement. The substituent of this metavariable can be provided either cataphorically by a subsequent demonstration by the interlocutors, as in (14.23), or anaphorically by appropriating a previous utterance event (perhaps not intended as a demonstration at all, as in (14.30)–(14.31)), or both, as in (14.24).

Finally, given that DS-TTR does *not* license form–meaning correspondences but only provides for the parsing/generation of stimuli in context, the process of reifying the demonstration of some linguistic element to provide ad hoc content of another already predicted type on the tree can be extended to the processing of non-linguistic signals composed with linguistic ones as the conceptualisation of some experience that is being demonstrated:

(14.35) John saw the spider and was like 'ahh!' [in a scared voice]

(14.36) John was eating like [gobbling gesture with hands and face]

Arguably, the modelling of the availability of such potential goes a long way in providing a starting point for modelling the processes required both for language acquisition and language change, both of which involve iterative, interactive implementation of this potential.

14.5 Conclusion

The view of NLs as codes mediating a mapping between internally specified expression types and private representations of the world has been abandoned here to give way to a view where utterances are taken as goal-directed physical actions coordinating with equally goal-directed mental actions within and across speakers. Instead of assembling propositional representations of reality, such actions are aimed to locally and incrementally alter the affordances of the internal/external context for both one's self and one's interlocutors so that the next predicted action goals can unfold. Taking the non-representational RT notion of procedures (Wilson 2016) as the basic architectural constituent of NL grammars, actions can be seen to manipulate perceptual stimuli composed not only of 'words' and 'syntax' but also of non-linguistic elements like visual marks, prosody, gestures, gaze, etc. NL meaning, from this point of view, is a

phenomenon that emerges at group level via interaction and is not determined through individual psychological processes. NL knowledge is then part of the ability to coordinate effective interaction with the environment, one's own self, or one's interlocutors. We believe that this non-modular, procedural definition of syntactic mechanisms removes one of the obstacles for integrating NL processing within a general sensorimotor account of higher cognitive functions (see e.g. Pulvermüller 2010; Gregoromichelaki et al. 2013b; Seth 2015; Pezzulo & Cisek 2016) as it avoids the necessity of the 'cognitive sandwich' (Hurley 2008) perspective. Furthermore, the necessity of assuming fine-grained interactional dynamics to account for meaningful engagements among agents and their environment indicates that the boundaries of individual cognition are fluid and indeterminate. Therefore, any account of individual psychological mechanisms will remain incomplete unless the distribution and complementarity of social cognitive processes are modelled in parallel (see also e.g. Bickhard 2009, 2016; Anderson 2014).

Acknowledgements

This chapter is part of an exploration of how basic RT insights could be extended to the concept of language itself, defining even syntax in terms of underspecification and incremental context-relative update. It comes with warmest thanks by the first author to Deirdre Wilson for her inspirational teaching and support, and by the second author to Deirdre for sustained friendship over many years. Grateful thanks are also due to two anonymous reviewers and the editors, in particular Robyn Carston, for valuable comments and advice.