

Procedural Syntax and Interactions

Eleni Gregoromichelaki
King's College London
University of Osnabrueck
elenigregor@gmail.com

Abstract

The view of NLs as codes mediating a mapping between “expressions” and the world is abandoned to give way to a view where utterances are seen as actions aimed to locally and incrementally alter the affordances of the context. Such actions employ perceptual stimuli composed not only of “words” and “syntax” but also elements like visual marks, gestures, sounds, etc. Any such stimuli can participate in the domain-general processes that constitute the “grammar”. The function of the grammar is dynamic categorisation of various perceptual inputs and their integration in the process of generating the next action steps. Given these assumptions, a challenge that arises is how to account for the reification of such processes as exemplified in apparent metarepresentational practices like quotation, reporting, citation etc. It is argued that even such phenomena can receive adequate and natural explanations through a grammar that allows for the ad hoc creation of occasion-specific content through reflexive mechanisms.

1 Language as action and grammar

Standard models that describe natural languages (NLs) as representational systems belong to the ‘language-as-product’ paradigm (Clark, 1992), concerned with the definition of linguistic representations, the “product” of linguistic processing. In this tradition, it has been a standard assumption that NL properties should be explained by reifying NLs as abstract codes, mapping forms (strings of symbols) to propositional intentions. However, a substantial amount of evidence indicates that NL use substantially affects NL structuring indicating an alternative characterisation: within a ‘language as action’ paradigm, NL properties can be explicated as coinciding with those of human action; an agent’s linguistic actions are structured sequentially, directed by predictions of upcoming inputs,

interleaved and interacting with other activities and agents. Accordingly, in everyday conversation, utterances are not expected to display evidence of necessary hierarchical constituency, e.g. sentential structuring: non-sentential utterances are adequate to underpin interlocutor coordination and all linguistic dependencies are resolvable across more than one turn:

- (1) Angus: But Domenica Cyril is an intelligent and entirely well-behaved dog who
Domenica: happens to smell
[radio play, 44 Scotland Street]

In such cases, postulating a notion of well-formedness based on a code licensing units ranging over strings of words, as an independent level of structuring, impedes a natural account of such phenomena. This is because joining overt forms together often results in illformedness or misleading interpretations:

- (2) A: I heard a bang. Did you hurt
B: myself? No, but Mary is in a state

Moreover, at the level of semantics/pragmatics of dialogue, the issue of recoverability of propositional intentions is also problematic, e.g., in cases such as (5) where various speech acts are accomplished within the unfolding of a shared single proposition (see Gregoromichelaki et al. (2011)):

- (3) Jack: I just returned
Kathy: from ...
Jack: Finland. [Lerner (2004)]
(4) Eleni: A: Are you left or
Yo: Right-handed. [natural data]
(5) Hester Collyer: It’s for me.
Mrs Elton the landlady: And Mr. Page?
Hester Collyer: is not my husband. But I’d rather you think of me as Mrs. Page. [The Deep Blue Sea (film)]

This endemic context-sensitivity and situatedness of NL use is indicative of the fact that both content and structure are emergent products of the

processes and practices underpinning human interaction. For these reasons, the more general approach to NL analysis argued for here revolves around the idea that structures, objects, concepts, concrete reality (and even the individual self) can all be taken as metaphysically emergent categories with processes, mechanisms, and change as ontologically primary.¹

2 DS-TTR

A grammar architecture adopting this perspective can be articulated within DS-TTR (Cann et al. (2005); Purver et al. (2010); Gregoromichelaki (in press)). Here NLs are conceived as comprising sets of processes modelled formally as *procedures*. Both NLs' temporal structuring (syntax) and lexical specifications are analysed as involving stored sequences (*macros*) of elementary (epistemic) actions, defined in an IF-THEN-ELSE format. Such actions incrementally and predictively build or linearise conceptual categories expressed in TTR-representations (Cooper, 2012). The model assumes tight interlinking of NL perception and action: production uses simulation and testing of parse states in order to license the generation of strings; comprehension predictively builds structures to accommodate upcoming inputs in order to constrain efficiently the usual overwhelming ambiguity of NL stimuli. By imposing top-down predictive and goal-directed processing at all comprehension and production stages, interlocutor feedback is incrementally anticipated and integrated. The model includes subsentential tracking of the shifting contextual parameters of each word-utterance event (Eshghi et al. (2015); Gregoromichelaki (in press)). *Context* constitutes an integral part of the grammar, not only as a record of the shifting parameters that provide for the interpretation of various indexical elements (e.g. *myself* in (2)), but also storing (a) the emergent (partial) structures constructed from the contributions of all participants; (b) the phonological/graphical elements that have been employed; (c) the actions used, recorded as traversals of paths in a graph display; (d) processing paths that have been considered as probabilistically live options but not eventually pursued (Sato, 2011; Hough, 2015). Storing the action paths is necessary

¹This view has its roots in an ancient philosophical programme starting in the Western world with Heraclitus, situated within a tradition following, among others, Martin Heidegger, Ilya Prigogine, Gilles Deleuze, and even encompassing current notions like the concept of the extended mind (Clark and Chalmers, 1998; Clark, 2008).

for the resolution of anaphora and ellipsis, especially “sloppy” or “paycheck” readings, whose resolution relies on re-executing (‘rerunning’) previous action sequences in an updated processing environment. Maintaining abandoned options is required for the modelling of backtracking in subsententially occurring conversational phenomena like clarification, self-/other-corrections, etc. but also humour effects and puns (Gregoromichelaki, in press). Consequently, coordination among interlocutors is seen not as inferential metarepresentational activity but as the outcome of the fact that the grammar consists of a set of licensed complementary actions that both speakers and hearers have to perform in synchrony (Gregoromichelaki et al., 2013).

2.1 Quotation in DS-TTR

Given these assumptions, a challenge that arises is how to account for the reification of grammatical processes as exemplified in apparent metarepresentational practices like quotation, reporting, citation etc. As we saw earlier in (1)-(5), perfectly intelligible moves in dialogue can be achieved simply by initiating a grammatical dependency which prompts either interlocutor to fulfill it without specific determination or identifiability of a given speech-act. In various other cases though, the interlocutor completing somebody else's utterance might be seen 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. (2)). There are further such phenomena in cases of citation, quotation, reports, echoing uses, and code-switching:

- (6) “Cities,” he said, “are a very high priority.”
- (7) Wright won't disclose how much the Nike deal is worth, saying only that “they treat **me** well”. [De Brabanter (2010)]
- (8) A doctor tells him [Gustave Flaubert] he is like a “vieille femme hystérique”; he agrees. [De Brabanter (2010)]
- (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. [Recanati (2010)]

Despite recent attempts to integrate such phenomena within standard grammars (e.g., (Ginzburg and Cooper, 2014; Maier, 2014; Potts, 2007)), certain data are not amenable to appropriate treatment due to the lack of modelling incrementality within these

formalisms. For example, as can be seen in (6)-(9), quotation can appear subsententially, and discontinuously, at any point, which means that contextual parameters regarding the utterance event and semantic evaluation need to be able to shift incrementally at each word-by-word processing stage. Additionally, quotation is one of the environments where the phenomenon of split-utterances is observed frequently as an opportunity arises for co-constructing a vivid unified perspective of some (actual or imaginary) speech/thought event (Gregoromichelaki, in press):

- (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)]

The contextual parameters relevant to the resolution of indexicals (e.g. *I*) in such cases, even though needing to shift mid-sentence, do not necessarily track the current speaker/hearer roles. Moreover, such role-switches include cases where the same structure can be employed both as expressing a speaker's own voice and as a subsequent quotation:

- (12) A: SOMEONE is keen [BBC]
B: says the man who slept here all night

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. For these reasons, an adequate account of the function of such NL devices can be given straightforwardly in DS-TTR 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 global propositional speech act (Gregoromichelaki (in press); Gregoromichelaki & Kempson (2016)).

On the other hand, modelling the potential of partially assuming another speaker's role, being perceived as "demonstrating" what somebody else was going to say, and the "metalinguistic" appearance of various such phenomena might seem especially problematic aspects for the DS-TTR stance:

- (13) "Life is difficult" is grammatical.
(14) James says that "Quine" wants to speak to us.
[James thinks that McPherson is Quine]
(15) "I talk better English than the both of
youse!" shouted Charles, thereby convincing
me that he didn't.

A DS-TTR grammar takes words (and the operation of "syntax" in general) as offering affordances exploited by the interlocutors to facilitate interaction. This means that words and linguistic constructions are NOT conceptualised as abstract code elements, expression types, that are associated with referential/semantic values (cf Cooper (2014) where string structure is still presumed). With no privileged semantic entities corresponding to (types of) expressions, only domain-general mechanisms for processing stimuli, quotation thus offers a crucial test for the legitimacy of these DS-TTR claims: when processing a quoted/cited string, what happens within the quotation marks (or any other indications) according to these assumptions?

In fact, it turns out that such cases are also unproblematic for the DS-TTR model, and can be explicated in a natural manner that conforms with intuitions and parallels the modelling of anaphora/ellipsis. First, in order to model cases like (6)-(10), (14), (15), as well as mid-sentence general code-switching, it has to be assumed that the context keeps track incrementally, through a designated metavariable (*g* in (16)), of which and whose grammar is being employed at each particular subsentential stage (cf Ginzburg and Cooper (2014)). Next, consider the most challenging cases, namely, metalinguistic uses, for example (13), so-called *pure quotation*, where an NL-string appears in a regular NP position. Under DS-TTR assumptions, this will be a pointer position where the grammar has already generated a prediction for the processing of a singular term (*?Ty(e)*, other cases might involve *?Ty(en)*, etc.). The explanation of what happens here is based on the fact that actions are first-class citizens in DS-TTR. This means that previous actions can be invoked by the grammar to be re-executed ('rerun') in order to provide parallel but distinct contents as needed in cases of sloppy-ellipsis or paycheck-pronoun readings. From this perspective, metalinguistic, echoic, and similar uses are cases where the actions specified by some grammar *g* for processing a particular string, e.g. the embedded sentential string in (13), come to be executed on the spot to provide an ad hoc conceptualisation of a demonstrated action. The formalisation of the basic mechanism is shown in (16) below. Different variants of this macro and combinations with other independently needed components of the grammar account for all such phenomena:

```

IF      ? $Ty(x_{\in\{e, cn, \dots\}})$ 
THEN   put  $Ty(x)$ 
          put  $(u_{q=\text{run}_g((a_i, \dots, a_{i+n}))} : e_s)$ 
(16) ELSE abort
      (a) demonstration action

```

In (16), the higher-order action `run`, also employed in cases of sloppy anaphora, is triggered. `run` is parameterised to some grammar g (replacing the metavariable \mathbf{g}), which can be distinct from the grammar used for parsing/producing the rest of the string (see (8), (15)). At the same time, the executed sequence of actions $\langle \alpha_i, \dots, \alpha_n \rangle$, bound to the rule-level variables $\langle \mathbf{a}_i, \dots, \mathbf{a}_n \rangle$, confers the ad hoc conceptual type of the quoting utterance event u_q which therefore functions as a demonstration. The performance of this demonstration event is then categorised as belonging to the already predicted semantic type, here, in (13), a referential term ($Ty(e)$) (feasible due to TTR’s subtyping definitions). The rest of the string then delivers a content that combines with the reification of this ad hoc execution. In (13), this delivers the interpretive result that this demonstration of the execution of the grammatical actions is characterised as having the property derived from processing *is grammatical*. For echoic cases, where the interpretation of the indexicals shifts following parameter values supplied by the invoked context, e.g. (7), (15), a similarly triggered action execution is accomplished, this time, with parallel introduction of the quoting context as a mentioned utterance event u , replacing the metavariable \mathbf{u} of type e_s , i.e. eventuality:

```

(17) IF      ? $Ty(x_{\in\{e, cn, \dots\}})$ , [ CONTEXT : [... [  $\mathbf{u} : e_s$  ] ] ]
THEN   put  $Ty(x)$ 
          put  $(u_{q=\text{run}_g}_{[\text{CONTEXT} : [\mathbf{u} : e_s]]} ((a_i, \dots, a_{i+n})) : e_s)$ 
ELSE   abort
      (b) demonstration-and-echoing action

```

Cases of direct quotation (e.g. (11), (12), (15)) are those where such a freely-available contextual switch has been grammaticalised in English.

Notably, given that the DS-TTR grammar does not provide form-meaning correspondences but only provides for the parsing/generation of stimuli in context, the same mechanism can be applied to non-linguistic signals/demonstrations: reifying the processing of some upcoming element to provide ad hoc content of another already predicted type explains how non-linguistic signals can compose sub-essentially with linguistic ones as the conceptualisation of some experience being demonstrated:

```

(18) John saw the spider and was like “ahh!”
(19) John was eating like [gobbling gesture]
(20) She went “Mm Mmmrn Mphh”

```

The existence of such compositions, along with all the previous data, might be challenging, under one construal, for the account of NL-gesture coordination in Rieser (2014; 2015). Rieser presents a framework (the λ - π calculus) where NL and gesture are modelled as independent but communicating processes. Even though the process metaphysics incidentally mentioned there is a welcome development, the assumption of independence might be questioned. First, this assumption seems to be an artifact of presupposing that NLS are structured codes mediating arbitrary mappings from standard syntactic forms (trees inhabited by words) to propositional meanings (e.g. λ -calculus formulae). Since the co-speech gestures examined are related to imagery (aural, visual, etc.) in an iconic manner, modelling their contribution in the standard way needs to abstract representations from the kinematics that cannot be unified with NL syntactic representations. In contrast, the view taken here is that NL actions do not require an independent syntax relying on the hierarchical structuring of stimuli sequences. Hence production/comprehension of stimuli in various modalities need not be segregated. Second, the major argument in Rieser’s analysis comes from SaGA data (Lücking et al., 2013) where NL segments and gesture-strokes seem not to synchronise perfectly. However, this is not an argument for considering such stimuli qualitatively distinct. Perfect synchronisation is not necessary within a single modality either, e.g. dialogue participants do not perfectly synchronise their turns. In a predictive framework like DS-TTR, such asynchrony might reveal a purpose, for example, co-speech gestures can be modelled as elaborating or narrowing down predictions that precede the processing of NL input. But then, under this view, there is a viable and useful application of the λ - π calculus in the DS-TTR framework too. Given that DS-TTR processing is strictly incremental pursuing only one path at a time, it is possible that various sources of information might compete for sequential positions. Introducing ad-hoc channel interfaces, modelled with resources from the λ - π calculus, can provide for the implementation of a sequentiality-repair mechanism, ordering inputs/outputs, even within the same modality, so that they can be processed strictly incrementally.

References

- Ronnie Cann, Ruth Kempson, and Lutz Marten. 2005. *The Dynamics of Language*. Elsevier, Oxford.
- Herman Cappelen and Ernie Lepore. 1997. Varieties of quotation. *Mind*, 106(423):429–450.
- Andy Clark and David Chalmers. 1998. The extended mind. *Analysis*, pages 7–19.
- Herbert H. Clark. 1992. *Arenas of Language Use*. University of Chicago Press.
- Andy Clark. 2008. *Supersizing the mind: Embodiment, action, and cognitive extension*. OUP USA.
- Robin Cooper. 2012. Type theory and semantics in flux. In R. Kempson, T. Fernando, and N. Asher, editors, *Handbook of the Philosophy of Linguistics*, volume 14: Philosophy of Linguistics, pages 271–323. Elsevier.
- Robin Cooper. 2014. Phrase structure rules as dialogue update rules. In V. Rieser and P. Muller, editors, *Proceedings of DialWatt - Semdial 2014: The 18th Workshop on the Semantics and Pragmatics of Dialogue, pages 4352, Edinburgh, 13 September 2014*.
- Philippe De Brabanter. 2010. The semantics and pragmatics of hybrid quotations. *Language and Linguistics Compass*, 4(2):107–120.
- Melissa C Duff, Julie A Hengst, Daniel Tranel, and Neal J Cohen. 2007. Talking across time: Using reported speech as a communicative resource in amnesia. *Aphasiology*, 21(6-8):702–716.
- Arash Eshghi, Christine Howes, Eleni Gregoromichelaki, Julian Hough, and Matthew Purver. 2015. Feedback in conversation as incremental semantic update. In *Proceedings of the 11th International Conference on Computational Semantics*, pages 261–271.
- Jonathan Ginzburg and Robin Cooper. 2014. Quotation via dialogical interaction. *Journal of Logic, Language and Information*, 23(3):287–311.
- Erving Goffman. 1979. Footing. *Semiotica*, 25(1-2):1–30.
- Eleni Gregoromichelaki and Ruth Kempson. 2016. Reporting, dialogue, and the role of grammar. In Alessandro Capone, Ferenc Kiefer, and Franco Lo Piparo, editors, *Indirect reports and pragmatics*, pages 115–150. Springer.
- Eleni Gregoromichelaki, Ruth Kempson, Matthew Purver, Gregory J. Mills, Ronnie Cann, Wilfried Meyer-Viol, and Patrick G. T. Healey. 2011. Incrementality and intention-recognition in utterance processing. *Dialogue and Discourse*, 2(1):199–233.
- E. Gregoromichelaki, R. Kempson, and C. Eshghi Howes. 2013. On making syntax dynamic: the challenge of compound utterances and the architecture of the grammar. In Ipke Wachsmuth, Jan de Ruiter, Petra Jaecks, and Stefan Kopp, editors, *Alignment in Communication: Towards a New Theory of Communication*. John Benjamins, “Advances in Interaction Studies”.
- Eleni Gregoromichelaki. in press. Quotation in dialogue. In & Johnson M. Saka, P., editor, *The Semantics and Pragmatics of Quotation*. Dordrecht: Springer.
- Julian Hough. 2015. *Modelling Incremental Self-Repair Processing in Dialogue*. Ph.D. thesis, Queen Mary University of London.
- Gene H. Lerner. 2004. Collaborative turn sequences. In *Conversation analysis: Studies from the first generation*, pages 225–256. Gene H. Lerner, John Benjamins.
- Andy Lücking, Kirsten Bergman, Florian Hahn, Stefan Kopp, and Hannes Rieser. 2013. Data-based analysis of speech and gesture: The Bielefeld Speech and Gesture Alignment Corpus (SaGA) and its applications. *Journal on Multimodal User Interfaces*, 7(1-2):5–18.
- Emar Maier. 2014. Mixed quotation: The grammar of apparently transparent opacity. *Semantics & pragmatics*, 7(7):1–67.
- Christopher Potts. 2007. The dimensions of quotation. In Chris Barker and Polly Jacobson, editors, *Proceedings from the workshop on direct compositionality*, pages 405–431. Oxford: Oxford University Press.
- Matthew Purver, Eleni Gregoromichelaki, Wilfried Meyer-Viol, and Ronnie Cann. 2010. Splitting the i’s and crossing the you’s: Context, speech acts and grammar. In P. Łupkowski and M. Purver, editors, *Proceedings of SemDial 2010 (PozDial)*, pages 43–50, Poznan, Poland, June 2010. Polish Society for Cognitive Science.
- François Recanati. 2010. *Truth-conditional pragmatics*. Clarendon Press Oxford.
- Hannes Rieser. 2014. Gesture and speech as autonomous communicating processes. MS, University of Bielefeld.
- Hannes Rieser. 2015. When hands talk to mouth. gesture and speech as autonomous communicating processes. In *Proceedings of SEMDIAL 2015-GoDIAL*, page 122.
- Yo Sato. 2011. Local ambiguity, search strategies and parsing in Dynamic Syntax. In R. Kempson, E. Gregoromichelaki, and C. Howes, editors, *The Dynamics of Lexical Interfaces*. CSLI Publications.