Actionism in syntax and semantics

Eleni Gregoromichelaki
Heinrich-Heine University and King's College London
Düsseldorf, Germany UK
elenigregor@gmail.com

Ruth Kempson
King's College London
UK
ruth.kempson@kcl.ac.uk

Christine Howes
University of Gothenburg
Sweden
christine.howes@gu.se

1 Introduction

In this paper, we present a view of "syntax" which is compatible with a perspective on perception called *actionism* (Noë, 2012). First we argue for the extension of the actionist view, which has been developed in the domain of low-level perception/action, to natural language (NL) on the basis that the motivating phenomena are parallel. We then show that the relevant NL phenomena include both semantic/pragmatic and syntactic issues and, on this basis, call for a dynamic conception of the 'grammar' that integrates both conceptualisation and syntactic licensing under uniform formal assumptions operating at the level of agent coordination rather than intra-individual mechanisms.

Actionism holds that perception is not a series of snapshots of scenes in the world leading to their inferential manipulation as representations in the brain as has standardly been assumed (Marr, 1982). Rather perception is engagement with the world, an activity and an achievement. The motivation for this perspective starts with the assumption that, in order to survive, organisms have to play an active part in controlling their environment and keeping it within desirable states. For an organism to exert such control, there must exist predictable relationships between its actions and ensuing perceptual stimulations (*sensorimotor contingencies*) since the purpose of perception/action is to ensure adaptability. Accordingly, any agent will benefit from actively exploring its material/social environment (its 'habitat', Heft, 1989) for risks or opportunities, with evolutionary processes ensuring that no heavy burden is placed on the cognitive resources required. Under this view, adaptive exploration and exploitation of environmental resources makes use of the agent's practical and embodied know-how of such sensorimotor contingencies, i.e., direct perception-action links (see, e.g, Buhrmann et al., 2013; Maye and Engel, 2011) rather than brain-internal cognitive inferential or representational means. Sensorimotor contingencies are lawful regularities in the dynamic relation between the agent and the habitat, patterns of dependence of changes in the sensory input as a function of an agent's movements (Gibson, 2014). Consequently, the information agents perceive about entities and their potential for interaction outcomes is agent-relative as it is mediated through the invocation of complex regular patterns, *constraints* (Barwise and Perry, 1983), originating from social as well as natural learning experiences. Various such learned expectations based on memorised holistic patterns of experience are built up through reiterated interactions with entities and are then deployed in subsequent encounters with them. But, at the same time, what the information agents perceive is also constitutively dependent on the niche they inhabit, the habitat, since information ensues only through their direct time-extended interactions with the sociocultural environment. "Perception" of an entity will then be constituted by the set of expectations it invokes concerning the possible interactions enabled through it (its *affordances*). This view is intended to replace the static, internalist-inferential view of "perception" as the association of stimuli with mental symbols stored and recovered as propositional knowledge.

2 Natural language as extended actionist perception

In our view, there are a number of parallels between the issues that the actionist view of perception aims to resolve and how NL comprehension (perception) and production (action) are inextricably and dynamically related both to the licensing of form and the construction of meaning.

2.1 Goal-directed contextual enrichment

The general problem that has led to internalist inferential theories is that perceptual understanding is not confined to what is immediately perceivable: it is generally agreed that the agent's perceptual capacities provide access to more than what is directly recorded on the stimulus or the presumed sense data. For example, in vision, we experience the total presence of features of the world, e.g., we see familiar objects as wholes, even though some of their parts or properties might be occluded. We encounter the same phenomenon in NLs in that we normally understand much more than what is explicitly encoded in an utterance:

(1) (a) Eleni: Leaving? (b) Frank: End of the month.

2.2 Goal-directed perceptual invariance

As the counterpart of this inevitable contextual enrichment, in object perception, we keep constant the experience of objects and their properties as they move through changing conditions. For example, we do not notice how the apparent colour of an object changes as we look at it moving from a bright environment outdoors to a less bright environment inside a building ('perceptual constancy'). Similarly, in NL use, speakers are usually unaware of the intricacies of the requisite syntactic/semantic coordination and the ambiguities and vagueness that decontextualised analyses of NLs present as problematic. For example, in dialogue, interlocutors frequently jointly develop a coherent single unit by skillfully continuing each other's turns while seamlessly adapting to subsentential local changes of contextual parameters (e.g. the referents and dependencies of indexicals) while observing other-initiated syntactic/semantic dependencies across turns and seamlessly shifting from one construal of a stimulus (*burn*) to another (the so-called phenomenon of "coercion"):

(2) [Context: A emerging from a smoking kitchen]
A: I've almost burnt the kitchen down.
B: Have **you** burnt
A: **Myself**? No... Well, only my hair.

2.3 Joint action as the source of normativity

For such cases in the domain of vision, actionism explains radical goal-dependence by emphasising the direct interdependence of perception and action: due to sensorimotor know-how, agents are capable of opportunistically pursuing affordances relevant to their current goals engaging with the habitat directly to confirm or disconfirm their expectations ('predictions') rather than aiming at the enrichment of intermediary brain-internal symbolic representations of the habitat prior to deciding on how to act to modify it. So the role of the brain's contribution is taken as a necessary but not sufficient factor in perception. Rather than orchestrating agent performance, the individual brain has considerable plasticity and capacity to support diverse and externally distributed behavioural repertoires. This is done through the temporary formation of nested and overlapping neural assemblies in which the same element can participate in various coalitions with other elements at different times (*neural reuse* Anderson, 2014) thus yielding distinct behavioural outcomes.

Generalising this view to NL, in any type of engagement with others or the environment, an agent acts in order to perceive the predicted consequences of its interactions instead of constructing and refining representations of these interactions to serve as guidance for its action. Such predictions are generated by means of the agent's embodied sensorimotor knowledge of the relevant habitat, i.e., by routinised expectations (the 'grammar') of how its various actions will change features of the sociomaterial world. But individual agent predictions are shaped and constrained by what is licensed within the current sociomaterial context, i.e., within the *normativity* of the socially-distributed nature of the grammar, so that no individual agent can be solipsistically aware of the significance of its own actions: by observing the consequences, the very act of speaking (or writing) in a particular context reveals to participants the normatively constrained triggers of actions for the words used as well as generating structured anticipations of further possible developments ('concepts'), the latter thereby becoming further affordances within that conversational exchange.

Given that normativity arises at the fluctuating sociomaterial level, such predictions inevitably and appropriately for adaptability (partially) fail. For this reason, NLs, as social objects, incorporate cultural practices that afford groups of agents online strategies for intervening and adjusting the landscape of affordances to the combined needs and goals of all the agents involved:

(3) (a) A: How would'ja like to go to a movie later on tonight?
(b) B: Huh?=
(c) A: **A movie** y'know like **a like ... a flick?**

(d) B: Yeah I uh know what a movie is (.8) **It**'s just **that=**

(e) A: **you don't know me well enough?** [from (Sacks, 1992)]

2.4 Concepts as active processes

This sensorimotor knowledge-as-action underpinning to cognition implicates conceptual understanding from the earliest stages of perceptual access (unlike existential phenomenology (Dreyfus, 2013) and related views). However, conceptual abilities do not, as in standard models, proceed via an intermediate cognitive stage before initiating the control of action, for cognition is not seen as separate from the sensorimotor grounding of agent performance. Under this view, concepts are not the rich internal representational structures of standard views – they are skills. It is argued that linking concepts exclusively to predicates in propositional judgements either in a direct (Kantian) way or an indirect (Fregean) way is inadequate from this perspective because there are other modes of activity where agents display conceptual abilities without propositional beliefs or judgements plausibly being involved (e.g., mundane everyday unreflective perception, reading in a familiar language, interacting with dogs, keeping appropriate social distances, etc.). For our purposes, we argue that in perceiving some entity and identifying it as a dog, it is not a static retinal image that becomes associated with the application of the 'DOG' concept. Instead, memorised patterns of current and past interactions are invoked to construct ad hoc a pattern of predicted interactions that differentiates the particular entity in the current context through its particular set of affordances as, e.g., a threat or a rewarding experience with incrementally adjustable behaviour of approach or avoidance (Gregoromichelaki et al., 2019; Bickhard and Richie, 1983). On this view, conceptual understanding cannot be taken as static pattern-matching but is, instead, an achievement. It is time-extended, incremental, and based on trial-and-error rather than an automatic mapping of experience to internal categories or propositional knowledge.

Moreover, due to their basis in action, concepts are necessarily always fragile and incomplete: in general, the specification of action guidance must allow flexibility to fit different situations and changing conditions and, therefore, successful situated action execution depends on leaving some degrees of freedom unbound (Suchman, 1987). This is notably echoed in NL phenomena like the so-called "polysemy" or "coercion" where word meanings are notoriously shiftable even within a single context (see, e.g., *burn* in example (2)).

2.5 The evolving nature of affordances

Both these degrees of freedom and the variety of multiple affordances in the human habitat introduce complexity due to the fact that agents do not perceive only one affordance at a time. Humans always perceive a continuously restructured dynamic field of affordances that consists of various possibilities for action soliciting attention. Cisek & Kalaska (2010) propose that 'affordance competition' is resolved by humans and animals through active moment-to-moment exploration of the field of available affordances without realising an overall plan of action but by being drawn towards the most rewarding predicted outcomes. Rietveld et al. (2018) have proposed that the "solicitation" of multiple complex affordances towards humans can be modelled as triggering states of 'action readiness' (Frijda et al., 2014). Perceiving (i.e. predicting) complex nested structures of potential affordances and developing appropriate action readiness requires training, developing skills, i.e., conceptualisations. For human agents, this is accomplished through participation in 'practices', i.e., coordinated patterns of behaviour of multiple individuals, within which NL interactivity is arguably the canonical case. Individuals or groups of individuals can then respond selectively to relevant (sets of) affordances in each particular situation because they act under the guidance of 'affective tensions', i.e., emotional responses like feelings of discontent or dissatisfaction, rather than "rational" deliberations through propositional beliefs/intentions. Such feelings of tension are aroused by the discrepancies (overwhelming prediction failure) between a concrete situation and the embodied skills of perceiving the norms of the situation type that the agent(s) have acquired by training. Agents resolve such tensions by resorting to their expertise. Their familiarity with the interactive environment allows them to intervene and restore perception of the expected affordances of the situation type. Again the NL case appears parallel, with, for example, practices of (non-sentential) clarification and correction in

(3b,c) or adjustment of expectations to differentiate a new situation type (e.g. proactively attempting to preempt social awkwardness in (3e).

3 NL grammar as (inter-)action coordination

To date, like the standard views of perception which actionism seeks to replace, formal theorising about NL has typically retained its characterisation as a code, an abstract system of rules and representations arbitrarily mapping forms to concepts conceived as symbols in a language of thought. On the view proposed here, to the contrary, NL is practice, underpinned by a set of conditional actions (the 'grammar') inducing ongoing continual flow of context, content, intentions, and speech acts. On this transformed view, NL is first and foremost coordinative action both with respect to the environment and other individuals; and a grammar formalism is duly defined directly in terms of defining the normative constraints (i.e. setting out and traversing the landscape of predicted affordances) that guide such action.

We take individual utterances as primarily physical events having effects (as stimuli) on human agents, both the utterer themselves and the perceiver (the addressee or any side-participants). Utterances can be further characterised as *actions*. Actions are physical movements realising goals (we include mental actions in this characterisation since, arguably, they are also realised by physical events within individual brains or social interactions). These goals are not formulated via the standard notions of (Neo-)Gricean intentions or plans but are, in fact, mostly, subpersonal, non-propositional, and unreflective, induced and resolved via the triggering of affective tensions and the employment expert know-how. As with perception, flexibility and efficiency requires that grammar-prescribed action specifications at various levels be partial so that the organism can adjust to its changing environmental circumstances. For example, efficient NL perception/production in dialogue is opportunistic at the subsentential level exploiting and exploring immediately what is made available by the interlocutor's local micro-actions:

(4) (a) Angus: But Domenica Cyril is an intelligent and entirely well-behaved dog **who**
(b) Domenica: **happens** to smell [BBC radio 4 play, 44 Scotland Street]

Of course, humans can form explicit goals and plans (propositional *intentions*), but even these have to be broken down into component subpersonal goals to be executed. Moreover, there is no one-to-one correspondence between a high-level intention and the implicit small-scale basic actions (mechanisms) employed to execute it. The reason is that the means employed to execute subgoals need to be responsive to what is available in the fluctuating context and this availability not only can modify explicit intentions, it is, in fact, the background for the generation of goals and intentions in the first place (Wittgenstein, 1953). So the Gricean notion of NL intention is derivative at best and arguably circular (Gregoromichelaki et al., 2011). Consider how an interlocutor can provide a grammatical context that prompts a speaker to expand their utterance just by fulfilling a pending grammatical dependency:

(5) (a) Jack: I just returned (b) Kathy: from
(c) Jack: Finland. [from (Lerner, 2004)]

Given that speakers are acting within a joint landscape of affordances and that normativity (i.e. goal success or failure) is defined at that social level, there is no need for explicit propositional declarations/inferences to the effect that joint action is maintained/failing (cf. Ginzburg, 2012). So, rather than having to figure out intentions, what is primitively available in the habitat (whether social or physical) are opportunities for action, corrective or advancing, i.e., *affordances*. Affordances which, under our interpretation are publicly available resources, trigger motivations for action within agents (*solicitations*). However, affordances are not, as standard, simply properties of the environment. Instead they are relations (Bruineberg et al., 2018) between agent abilities and what the current sociomaterial environment reliably makes available. This means that the shifting set of affordances in dialogue concerns the collective potential of the interactants, rather than individual perspectives whose meshing needs to be explicitly negotiated/represented. Instead, the local and shifting landscape of affordances provides for a joint conceptualisation of the current action potential with minute adjustments at each subsentential stage resulting in the appearance of planned rational action at the macrolevel:

(6) A: so . . . umm this afternoon . . .
B: let's go watch a film

A: yeah

(7) (a) A: I'm pretty sure that **the**:
(b) B: **programmed visits**?
(c) A: programmed visits, yes, I think they'll have been debt inspections. [BNC]

As Gibson (2014) suggested, humans and animals perceive the world in terms of affordances rather than in terms of low-level objective features of the environment. For us, this means not only that we do not perceive the world in terms of the categories studied in physics (molecules, atoms, etc.) but also not in terms of individuated descriptive concepts like the atomic symbols of a language of thought. We extend this view to NLs, assuming that grammars provide direct access to, or means of intervention in, the conceptual articulation (the affordances) of the sociomaterial human habitat. Consider, for example, how the use of a single accusative-marked DP in Greek characterises an agent's action as incompatible with some selected property of an entity in the visual environment:

(8) [Context: A contemplates the space under the mirror while re-arranging the furniture; B brings her a chair]
A: tin karekla tis mamas? / #i karekla tis mamas?
the$_{acc}$ chair$_{acc}$ of mum's? / #the$_{nom}$ chair$_{nom}$ of mum's?
(Ise treli?) (Are you crazy?) [Modern Greek]

The utterance with the accusative marker allows the differentiation of the entity (the chair) as the inappropriate 'Patient' of some unspecified action by the listener, the latter aimed to be compatible with the current joint goals. Given these joint goals, linguistic and physical actions mesh directly with each other and their interleaving eliminates the need to resort to propositional or syntactic expansions of non-sentential utterances (NSUs).

Moreover, unlike the standard view claiming that we decide what to say (cognition) before specifying how to say it (action), we argue that NL action selection happens during the continuous micro-interaction with the world/interlocutor, without representation of other agents' psychological states and knowledge. As can be seen in the examples earlier (e.g. (6)) and below in (9), we do not need to assume that speakers plan whole propositions or speech acts before they can start speaking. Instead, interlocutors can rely on each other for action completion (6) and are, through their coordinated activities, able to locally adjust their language, their relationships, and the environment to fit the fluctuating circumstances:

(9) Tess: Okay, so we were not exactly invited. But he's here, and we're here, so that makes us ...
Jack: total idiots!
Tess: in the right place at the right time.

Given this perspective, our dynamic approach to NL maintains that what is important for grammar modelling is the time-involving and interactive properties of an NL system while, given data from everyday joint activities, no representational, metalinguistic notion of "complete sentence", or even "syntactic constituent", is required for explaining NLs. Such constructs are not notions that are fundamentally part of the awareness employed in everyday NL use and, for this reason, we argue, theoretically redundant beyond the analysis of written or preplanned discourses. (Linell, 2005; Gregoromichelaki et al., 2009, 2011; Kempson et al., 2016, 2017). In fact, such notions impede natural characterisations of how NL elements contribute to the achievement of agent coordination. As can be seen in (1), (8), it is clear that NSUs are adequate in context to underpin conversational interaction making complete and efficient contributions. As they mesh seamlessly with people's physical activities, public (re)employment and negotiation of the affordances of any NL signal shifts attention towards selected aspects of the current experience (*conceptualisation*) so that various *joint-projects* (Clark, 1996) can be pursued. Such joint-projects (or *language-games* Eshghi and Lemon, 2014; Eshghi et al., 2017) can then be advanced just by use of even minimal NL contributions (e.g., *huh?* in (3b)), gestures, eye gaze, and emotional displays, without any need to characterise such functional stimuli as in any sense "elliptical" and in need of syntactic/propositional expansion.

Given the methodology of modelling incrementality, any lexical action can be seen, on the one hand, as potentially complete, having effects in its own right but, also, as a trigger for further processing (a *constraint*) by being perceived as embedded within a wider action context. In this way, the local adaptive dynamics of co-action impose an overall structuring in language-games of various scales under which role differentiation and joint responsibility (*action complementarity*) can be induced and sustained without explicit cognitive/public representations of what the agents seek to accomplish (Mills and Gregoromichelaki, 2010). For example, agents – just by assuming incremental processing – can induce their inter-

locutor to provide the input required to complete their own actions, thus actualising ad hoc the performance of what have been described as conventional *adjacency pairs* or speech acts (see also earlier, e.g., (5) (Gregoromichelaki et al., 2013):

(10) (a) Psychologist: And you left your husband because
(b) Client: we had nothing in common anymore

(11) (a) Jane: u:m Professor Worth **said that**, if Miss Pink runs into difficulties, on Monday afternoon, with the standing subcommittee, over the item on Miss Panoff,
(b) Kate: **Miss Panoff?**
(c) Jane: yes, **that Professor Worth would be with Mr Miles all afternoon**, - so she only had to go round and collect him if she needed him [from (Clark, 1996): 240-241]

As can be seen from all the examples above, given that the grammar is a set of constraints underpinning joint action, any type of syntactic/semantic dependency can be set up and resolved across more than one turn with the resolving element satisfying expectations generated by the utterance of either interlocutor. Moreover, by shifting the focus of NL analysis away from the presumed denotational/referential function of NL strings to their procedural and dynamic potential, we can observe that initiation of what have been characterised as purely syntactic dependencies can operate as ad hoc speech-act indicators, i.e., newly-introduced affordances to prompt the interlocutor to act.

3.1 Syntax/morphology as constraints on affordance fields

Shifting the view of syntax away from representations to a set of procedures complementary to all other actions in dialogue does not mean that we deny its significance. Even though complete sentences/clauses are not necessary in dialogue processing, morphosyntactic constraints are implicated in the incremental continuity of discourse and the choice and licensing of NSUs as already shown earlier in (8). Additionally, in English and other languages, the obligatory binding of a reflexive pronoun can be distributed over turns uttered by distinct interlocutors shifting its form in accordance with contextual parameters that subsententially switch as they track the current speaker/addressee roles (see (2) earlier). Moreover, in morphologically-rich languages, speech acts with subpropositional elements, e.g., requests as in (12) below, and interjections as in (8), require the presence of appropriate 'agreement' morphemes, e.g. case, gender, number, indicating how the uttered "fragment" will induce selection of pertinent affordances from the context created by the utterance:

(12) [Context: A goes into a coffee shop to order coffee]
A to B: (ena) metrio me gala /
(a-$_{acc-masc-sing}$) medium$_{acc-masc-sing}$ with milk
#metries me gala
#medium$_{acc-fem-pl}$ with milk
(A) medium (-sweet coffee) with milk
[request, Modern Greek]

This shows that, rather than inference being required to enrich NSUs to propositional/sentential forms, morphosyntactic constraints play an active role in affordance competition by directing attention to the relevant aspects of the situation. For example, in (12), the accusative-singular-masculine marking on the adjective ('moderate(ly-sweet)') just narrows down the already present set of affordances of the environment (a cafe) by identifying the relevant properties of the 'Goal' involved in the speaker's action. We do not have to assume that some propositional representation needs to be constructed to fit in the "fragment's" contribution. Such morphosyntactic constraints are not empty, arbitrary, and/or parasitic on some primary referential function. Instead, they are used as conceptual resources to differentiate, ad hoc (in (8)), or within more socially established behavioural settings (Heft, 1989) in (12), a salient set of situated affordances which impromptu constitute the entity involved. Accordingly, physical and grammatical NL actions readily compose with each other exactly because they perform meshing contributions in human interaction (Gregoromichelaki, 2017):

(13) She played [PLAYING TUNE ON THE PIANO] not [PLAYING ANOTHER TUNE ON THE PIANO]

(14) OK, let's do it together. So we have [ARM MOVEMENT DEMONSTRATION] and then we go [LEG MOVEMENT DEMONSTRATION]

3.2 Incremental prediction

Under this view of NL syntax and content, incrementality means, first, that during production, interlocutors do not need to plan whole propositional units before they start speaking. Instead, they need to generate multiple local (probabilistically ranked) predictions of the following perceptual inputs (multimodal stimuli or the other agents' active feedback) for themselves and the interlocutors. This means that they always anticipate how their projected units (words, phrases, or non-NL-actions) will affect the context, which includes the

other interlocutors' reactions and changes of their own perceptual stimuli. Through the subsequent process of *affordance competition*, producers can then select a minimal NL action that would ensue as the most rewarding short-term outcome concerning the (joint) task (see Cisek and Kalaska, 2010). This is why speakers can unproblematically integrate gradual modifications of their utterance (e.g. repairs, new interlocutors entering the scene, etc) induced either by themselves (3c) or their interlocutor (4)-(11); and they can go on extending and elaborating either their own utterance (11a) or the one offered by an interlocutor (7c). Thus, the production process is very tightly incrementally coordinated with the interlocutors' responses as they come because it includes a fine-grained incremental feedback loop that controls and procedurally coordinates all participants' actions (Goodwin, 1981; Bavelas et al., 2000).

Secondly, during comprehension, in the same way, efficient incremental procedural coordination demands that addressees also continuously predict a range of upcoming stimuli and check whether the actions of their interlocutor and actually perceived stimuli conform to those. Thus listeners/perceivers incrementally generate and seek the satisfaction of a range of local predictions, intervening in a timely manner where their anticipations are found in over-threshold error and some "surprising" input cannot be integrated as an unforeseen but adequately rewarding outcome (see, e.g., (6) vs (9)). This local adjustment to task requirements via affordance competition avoids the need to impose the necessary calculation of whole propositional intentions or even implicate (an infinite regress of) mutually known facts. Experimental and empirical conversation analysis (CA) evidence shows that interlocutors do not engage in complex mind-reading processes trying to figure out "speaker meaning", neither do they even need to calculate common ground (Engelhardt et al., 2006, a.o.). The reason for this is that each agent during an interaction does not act independently to realise a predefined action plan, in fact, often, no such plan exists or only emerges post hoc – independently of the agents' explicit goals (hence the value of conversation).

As a result, given incremental processing, interlocutors can abandon unfruitful courses of action midway (see (3c)), even within a single proposition, without, nevertheless, presupposing that such productions will be taken as having remained unprocessed:

(15) A: **Bill**$_i$, who . . . , sorry, Jill, **he**$_i$'s abroad, she said to let me finalise the purchase.

This leads to a rather different perspective on such "repairs". Even though useful as a descriptive characterisation of normative practices (Schegloff, 2007), singling out a notion of "repair" for explicating the function of such NSUs is misleading: from a dynamic modelling perspective, any behaviour in dialogue is already taken as aiming to control perception (feedback), with perception in turn providing motivation for adjustments via further action. In a sufficiently fine-grained dynamic model, repair as a separate category of constructions (Clark, 1996) turns out to be an artifact of assuming that the interlocutors aim for the establishment of shared common world "representations" employing speech acts that contribute propositional contents (Poesio and Rieser, 2010; Ginzburg, 2012) in the service of reasoning and planning. Instead, we can see the goal of feedback control, striving to integrate 'prediction error' (Clark, 2017a,b), as a constant local aim and structuring factor of any (joint) activities.

There are complementary pressures here, as on any group activity. From the intra-individual psychological point of view, it is the mechanisms of processing NL signals which invoke selective aspects of previous experience with such stimuli ('solicitations'), while inter-individual feedback leads to the ad hoc creation of temporary inter-individually distributed "grammars" and "conceptual structuring" (in the Wittgensteinian notion of "grammar" (Wittgenstein, 2005), for us, the local 'field of affordances'). Thus, concepts, like words, are just the triggers of further action-organising affordances inducing the prediction of further possible outcomes in the form of anticipated feedback from the interlocutor or the environment (see also (Cisek and Kalaska, 2010)). These second-order affordances need to be incrementally reconstructed (enacted) each time. But, with repeated use, conceptual mechanisms, like syntactic (sequence-processing) mechanisms, establish gradually reinforced memory traces that pick up encapsulated easily recoverable nested sensorimotor routines (*macros*, i.e., complex constraints). Therefore conceptual mechanisms are also part of the grammar and can be seen as relatively entrenched, culturally-enabled abilities to track cul-

turally or environmentally significant invariances (Millikan, 2005; Casasanto and Lupyan., 2015). Processing words and syntactic structures, like other stimuli, trigger these processes of conceptualisation, and participants in a dialogue need to coordinate on these procedures as well as their physical actions (e.g. turn-taking).

Taken together, these empirical facts show that physical action, syntactic licensing, and conceptual processing are performed incrementally subsententially and in tandem, underpinned by the same mechanisms, and, at each step, affording possibilities for further extension by the interlocutors' actions or the situational context. Giving due recognition to the foundational nature of dynamic practices of interaction, as we shall now see, we can ground the appearance of presumed phenomena of "conventionalisation", "processing economy" (Kirby, 1999; Carston, 2002) or "signal economy" (Langacker, 1977) – all exemplified by NSUs – in the plastic mechanisms of action coordination rather than burdening inference or representational computation. But this requires viewing NLs as skills implemented by domain-general procedures rather than fixed form-meaning mappings. And we now turn to providing a sketch of a procedural grammar architecture whose explicit aim is to directly model such a conception of NLs.

4 Dynamic Syntax: Language as action

4.1 Syntax as state transitions

Dynamic Syntax (DS, Cann et al., 2005; Kempson et al., 2001) is a grammar architecture whose core notion is incremental interpretation of word-sequences (comprehension/perception) or linearisation of contents (production/action) relative to a temporally fine-grained notion of context. The DS syntactic engine, including the lexicon, is underpinned by a specialised version of Propositional Dynamic Logic (PDL), which is a formalism able to express probabilistically licensed transition events among the states of a dynamic system (Sato, 2011). As a result, DS is articulated in terms of conditional and goal-driven actions whose accomplishment either gives rise to expectations of further actions, tests the environment for further contextual input, or leads to abandonment of the current strategy due to its being unviable in view of more competitive alternatives. Words, morphology, and syntax are all modelled as "affordances", i.e., indicators of opportunities for (inter-) action. Such interactions incrementally open up a range of options for the interlocutors so that selected alternatives can be pursued either successfully or unsuccessfully: even though a processing path might look highly favoured initially, due to the changing conditions downstream, it might lead to failure so that processing is aborted and backtracking to an earlier state is required (Sato, 2011). The potential for failure or success relative to goals imbues the activities of the system, even though mainly subpersonal, with a notion of normativity arising from the routinisation of action sequences retrievable as chunks (*macros*). Such macros impose licensed expectations (predictions) that can in turn operate as triggers resulting in nested structures of affordances constraining potential interactions. This normative field of nested anticipations of further interactions built on the basis of prior trial-and-error efforts comes to constitute an instantiation of the *grammar* in particular concrete occasions. Such ad hoc grammars are what prompts or constrains the actions of the individuals participating in a dialogue. Following the opportunities opening up by their recognition of affordances (or avoiding paths that might lead to trouble), interlocutors perform step-by-step a coordinated mapping from perceivable stimuli (phonological strings) to conceptual and physical actions or vice-versa.

To illustrate, we display in Fig 1 the (condensed) steps involved in the parsing of a standard long-distance dependency, *Who hugged Mary?*.[1] The task starts with a set of probabilistically-weighted predicted *interaction-control states* (ICSs) represented in a directed acyclic graph (DAG). At this stage, let's assume the first utterance in a dialogue, the DAG landscape displays all the potential opportunities for parsing or producing relative to the habitat, prompting lexical actions as licensed by the grammar of English. These potential actions are assumed to be "virtually present" for the participants even though they are not all eventually actualised.[2] Either participant might take the initiative to begin the articulation of an utterance while the other is in a state of preparedness checking

[1] The detailed justification of DS as a grammar formalism is given elsewhere (Kempson et al., 2001, 2011, 2016, 2017; Eshghi et al., 2011, a.o.).

[2] For relevant notions of "virtual presence", see (Noë, 2012; DeLanda, 2013)

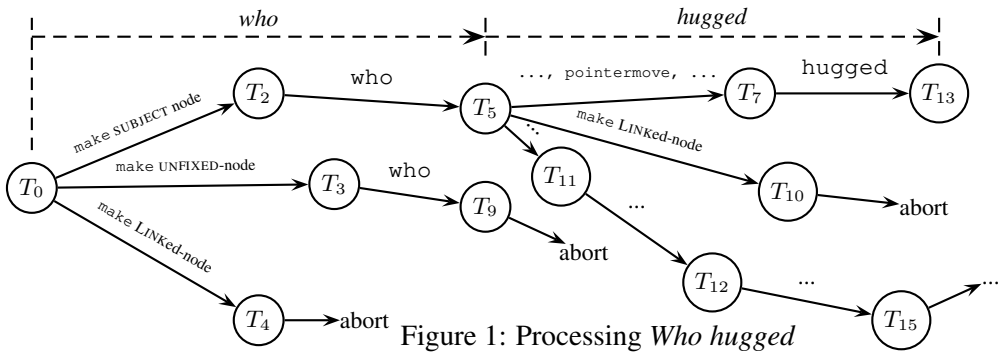
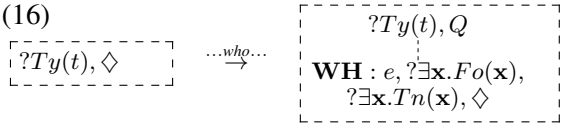


Figure 1: Processing *Who hugged*

whether the path pursued by the other interlocutor conforms to their expectations or whether they need to take over and compensate for their lack of coordination (Eshghi et al., 2015). Many alternative processing paths unfold at each step as affordances of the environment and the interlocutor are taken up or are gradually abandoned (see also Sato, 2011; Eshghi et al., 2013; Hough, 2015).[3] An ICS field tracks the conceptualisation of salient habitat information, implements means of coordination, e.g. backchannels and repair (Eshghi et al., 2015; Howes and Eshghi, 2017), and records the recent and projected history of processing. On this basis, each ICS node contains an indicator of the current focus of attention, the *pointer*, $\Diamond$, which is crucial for the time-linear unfolding of processing as its various positions define distinct potential developments. As far as NL signals are concerned, the pointer is responsible for word-order regularities in any particular language so that processing is constrained with respect to its potential continuations. Since each ICS node includes a pointer position, it will induce a specific cascade of grammatical goals (*requirements*) to build/linearise conceptual structures ('ad-hoc concepts') constrained by what is made available by the macros that constitute the practical knowledge of the language.

Individual NLs impose a particular conceptualisation of states-of-affairs given what is available in its lexicon and morphological resources. For example, in English, the verb *disappear* only requires a subject whereas the corresponding verb in Greek requires an object as well.[4] Therefore, the conceptualisation affordances in each NL are distinct and the expectations for further perceptual input or action induced at each ICS need to be in accordance with what can be formulated in that NL. For this reason, building language-appropriate conceptualisations is guided in DS by labels characterising ontological types (e for entities in general, e_s for events, $(e \rightarrow (e_s \rightarrow t))$ for one-place predicates ('disappear', in English), $(e \rightarrow (e \rightarrow (e_s \rightarrow t)))$ for two-place predicates ('disappear' in Greek), etc.). In (16) below, focussing now on only one snapshot of an active DAG path in Fig 1 (and only the syntactically-relevant part), we see that the initial goal (indicated by ?), in this case, happens to be realised as a prediction to produce/parse a proposition of type t. Below, on the left, this is shown as a one-node tree with the requirement $?Ty(t)$ and the ICS's current focus of attention, the pointer $\Diamond$:

(16)



Such predictions can be satisfied either through processing a stimulus produced by an interlocutor, by attending to a stimulus from the physical environment or by the agent themselves producing the requisite mental or physical actions that fulfil the predicted goal. If linguistic satisfaction of the goal is chosen, either through an interlocutor or the self, as shown in (16), the pointer at a node including a predicted type t outcome ($?Ty(t)$) will drive the generation of further predicted affordances/subgoals. In this particular DAG path, preparation needs to be made for accommodating the processing of the lexical stimulus *who* whose affordances are expected to be part of the eventual satisfaction of the current $?Ty(t)$ goal.

In (16), one of the probabilistically highly-favoured next steps for questions in English is

[3] A more realistic graph would also include the possibilities of non-verbal actions, not only gestures, but also physical voluntary actions like, for example, the physical response to a command or request. It is our claim that any "speech act" can be performed non-verbally (see, e.g., Clark, 2012 and earlier (13)-(14)).

[4] O Giannis exafanise *(to vaso).
The Giannis disappeared *(the vase).
John caused the vase to disappear.

displayed in the second partial tree: a prediction that a structurally underspecified node (indicated by the dotted line) can be built and can accommodate the result of parsing/generating *who* along with an indication of interrogative mood (Q). This reflects the fact that for speakers of English, perceiving *who* sentence-initially is constituted by realising affordances of introducing expectations for a *wh*-question coming up (among other potential). According to DS, realisation of these further affordances for English will be achieved by a combination of executing both lexical and general tree-building action macros that are conditional on certain contextual factors being present (e.g., this being the first word uttered in the sentence) and, in turn, imposing new goals for further processing. For example, given the impoverished nature of case-marking in English, as illustrated here, temporary uncertainty about the eventual contribution of an element like *who* (subject vs object, etc.) is implemented through *structural underspecification* accompanied with an expectation ($?\exists \mathbf{x}.Tn(\mathbf{x})$) that further processing will resolve the uncertainty. Initially so-called "unfixed" tree-nodes model the retention of the contribution of the *wh*-element in a memory buffer until it can be used. Further processing is expected to yield a situation where an argument node is required and no lexical action is provided so that the unfixed node can then be retrieved to satisfy the goal of achieving a licensed tree substructure within the local tree domain. Moreover, grammatical words like *who* and other semantically weak elements (e.g. pronominals, anaphors, auxiliaries, tenses) contribute radically underspecified content in the form of so-called *metavariables* (indicated in bold font), which trigger search ($?\exists \mathbf{x}.Fo(\mathbf{x})$) for their eventual type-compatible substitution from among contextually-salient entities or predicates.

General computational and lexically-triggered macros then intersperse to develop a binary tree: in Fig. (2), the verb *hugged* is next processed. It contributes conceptual structure in the form of unfolding the tree further and assembling an ad-hoc concept (indicated as Hug') developed according to contextual restrictions,[5] It also introduces placeholder metavariables for time and event specifications ($\mathbf{S}_{PAST} : e_s$) whose values need to be supplied by the non-linguistic affordances of the current ICS.

4.2 Conceptualisation as state transitions

The conceptual structure being built here is indefinitely extendible (see Cooper, 2012) and "non-reconstructive" in the sense that it is not meant as a passive inner model of the world (see also Clark, 2017a,b) but as a means of interaction with the world via the predictions generated regarding subsequent processing. Accordingly, the affordances that constitute the conceptual structure are viewed as relational (see also Chemero, 2009; Bruineberg et al., 2018): a pairing of (aspects of) the world with a (joint) perspective, namely, those affordances of the sociomaterial world that are accessible relative to the agent(s)' relevant sensorimotor skills shaped by prior experiences and the econiche.[6] Here the perspectival construal of types, as accessible affordances/constraints, permeates the very definition of what an affordance is. It is, therefore, a feature that is constantly present in what agents perceive/achieve. Following standard assumptions in ecological psychology and phenomenology, it is part of the force of an affordance that the perceiving/acting agent becomes aware that they are manipulating the world from a particular point of view. This awareness is enabled as part of the agent's sensorimotor knowledge of regularities and lawful variations regarding the changes in the environment that are caused by the agent's own actions as opposed to actions/events affecting the agent. As a result, when multiple agents are coupled as a temporarily formed agentive system, or in cases where experts use tools or patients use prostheses, the collective perception/action possibilities that emerge for the newly-formed unit are not the result of simple summation of what is possible for the individual components. The joint landscape of affordances can be much more or much less depending on "enabling" or "disabling" couplings. In both cases, agents are able to perceive this new regime and generally capable to adjust their contributions in complementary ways (Mills and Gregoromichelaki, 2010; Mills, 2014).

The relativisation of the structure of human con-

[5] In Purver et al. (2010), this is modelled as a *record type* via a mapping onto a Type Theory with Records formulation, but we suppress these details here: see Purver et al. (2011); Eshghi et al. (2013); Hough (2015); Hough and Purver (2014); Gregoromichelaki (2017); Gregoromichelaki and Kempson (2018).

[6] In this actionist and externalist perspective, we diverge from standard construals of TTR as in (Ginzburg, 2012), Cooper, forthcoming.

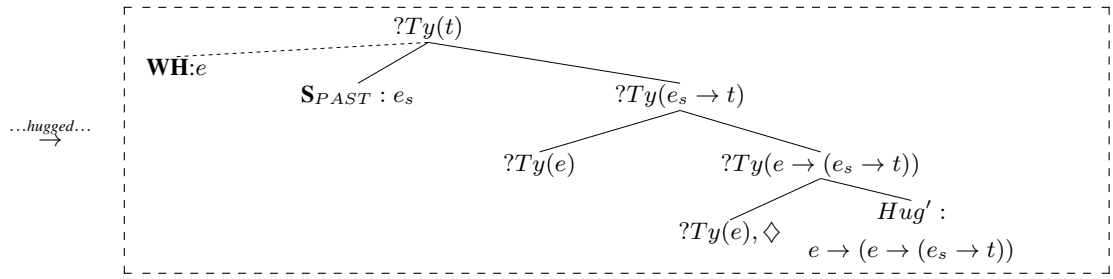


Figure 2: Processing *hugged*

ceptual types against practice-based abilities has normative implications, in that the agent(s) might fail to achieve what is genuinely afforded to them by the sociomaterial environment, or the agent(s) might fail to take up the multitude of affordances that have been perceived as potential ("virtual") paths of action. Moreover, given that they engage with real properties of the sociomaterial habitat, the consequences of misapplying their abilities will be detectable by the agents themselves as error signals when their predictions are falsified. Such failure is the source that can lead to repair and adjustment so that long-term learning and adaptation are the outcomes.

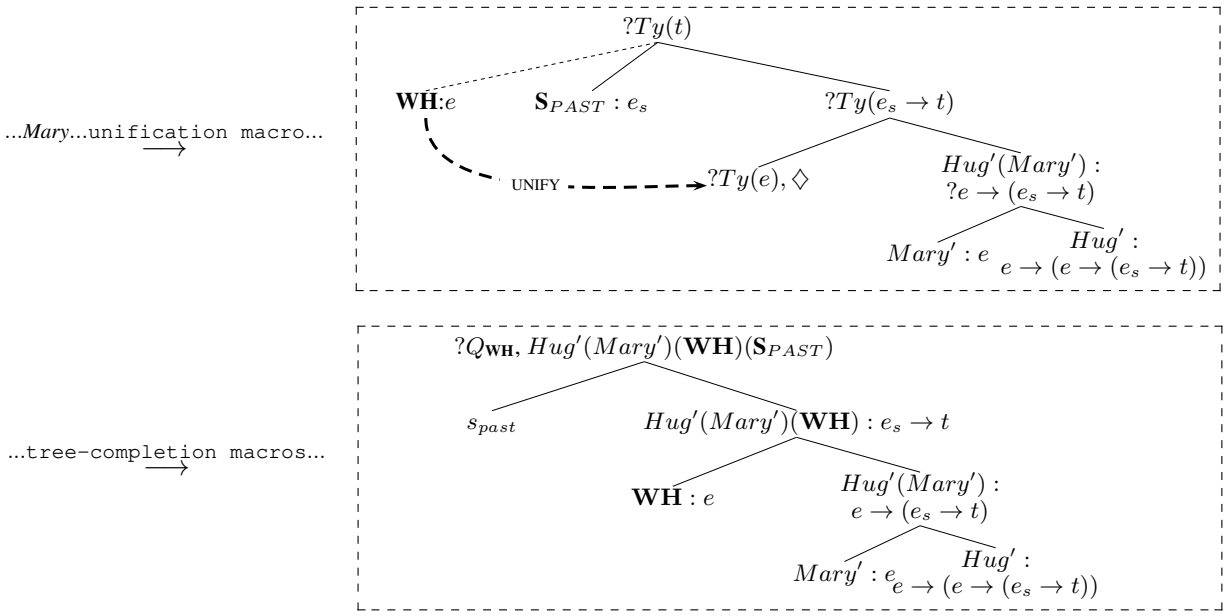
Given the requisite dynamicity and world grounding, concept labels, like Hug' here stand for abbreviations of triggers for complex sets of action potentials embedded under the DAG nodes as nested affordances. Such labels then constitute additional ICS choice points in the generation of further potential paths within the DAG. Given this view of concepts, what individuates each such label is their distinguished provision of sets of available actions realisable in the next steps within the affordances field (the DAG). Since we take perception and NL-comprehension as a time-extended and incremental activity, the manifestation or awareness of such a concept will develop gradually rather than instantaneously in an act of judgement. To take a "syntactic" type as example, type t is differentiated from type $(e_s \rightarrow t)$ in that the former (minimally) leads to the prediction of a left daughter of type e_s and a right daughter of type $(e_s \rightarrow t)$ whereas the latter leads to the prediction of e and $(e \rightarrow (e_s \rightarrow t))$. This is what differentiates these types not their distinct labels. Within the grammar, such types either contribute tests in the conditional procedures that implement the operation of grammatical and extra-linguistic actions or trigger searches for appropriate words, or expand the current structure and annotations with the anticipation of further developments. Even more pertinently, they do not have any model-theoretic content beyond the transitions they allow or curtail in the traversal of the states of the PDL model that underpins DS. Similarly, we take concept labels such as Hug' as triggering access to nested structures of potential actions regarding aspects of (mental or physical) interaction with an event of hugging, some of which will be taken up and others abandoned. As such, the types (concepts) are mainly constituted by subpersonal mechanisms, however, the results of their operation can be brought to consciousness by processes of reification for purposes of, e.g., linguistic negotiation, explicit planning, theory construction, or teaching.

Given affordance competition, agents select their next actions based on possibilities (probabilistically) grounded on these types which function as 'outcome indicators' (Bickhard and Richie, 1983) so that the predictions yielded by these types might be reinforced (verified) or abandoned (fail) in the next steps. As long as they remain as live possibilities, the operations induced by the types will keep triggering flows of predictions for further (mental or physical) action even if particular paths of sequences of nested predictions are not taken up. Maintaining even abandoned options is required for the explicit modelling of conversational phenomena like clarification, self/other-corrections, etc. but also, quotation, code-switching, humorous effects and puns (Hough, 2015; Gregoromichelaki, 2017):

(17) John went swimming with Mary, um. . . , or rather, surfing, yesterday.
['John went surfing with Mary yesterday']

(18) The restaurant said it served meals any time so I ordered breakfast during the Renaissance.
[Stephen Wright joke]

So, the contribution of the verb *hug* to the DAG would be a conceptual type here just labelled as



Hug' to encompass the set of relevant affordances that are predicted as potential further engagements with an event of hugging. As part of its "syntactic" contribution, which we do not consider as qualitatively distinct given what we discussed earlier with respect to *disappear* in Greek and English, *hug* will also introduce the prediction of an upcoming invocation of an entity that undergoes the hugging action (the 'Patient' role). This is implemented by the construction of a new node on the tree in order to accommodate this predicted occurrence. Now returning to the processing stage displayed in Fig (2), we see that the pointer $\Diamond$ is residing at this predicted argument node ($?Ty(e)$). This implements the word-order restriction in English that the object needs to follow the verb. In NLs with morphological cases, like Greek as seen in (8), (12) earlier, it will be the inevitable case morphology instead that induces narrowing down the available properties of the noun content to fit a particular role assignment ('Patient') in some event conceptualisation triggered by a verb or the physical situation. For this reason, DPs in Greek can appear in a variety of positions in the sentence and they place much less requirements for contextual support than in English where the thematic role is not immediately predictable.

Returning to English now, at the stage shown in Fig. (2), the word *Mary* can be processed to initiate the tracking of a contextually-identifiable individual ($Mary'$) at the argument node internal to the predicate.[7] 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. The presence of the metavariable on this node eventually results in an ICS that includes a requirement for the provision of a value for the metavariable, in effect an answer to the question posed by the utterance of *Who hugged Mary?*, imposed as a goal ($?Q_{\mathbf{WH}}$) for the next action steps (to be resolved either by the speaker or the hearer), see Fig. 4.2

4.3 Coordinating comprehension-production

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. *Production* equally involves the generation of predictions/goals but, this time, also the deployment of action (verbalising) by the predictor themselves in order to accomplish their predicted goals. By imposing top-down predictive and goal-directed processing at all comprehension/production stages, interlocutor feedback or changing of direction due to perceiving one's own action consequences ('monitoring') is constantly anticipated and seamlessly integrated in the ICS (Gargett et al., 2008, 2009; Gregoromichelaki et al., 2009; Purver et al., 2010; Eshghi et al., 2015). Feedback can ex-

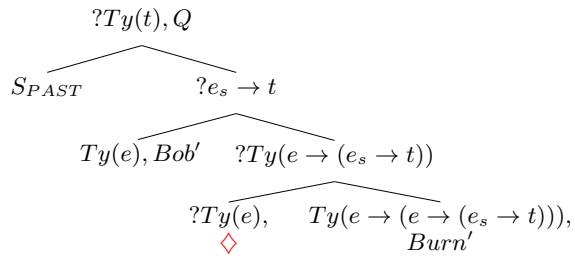
[7] For the view that such entity concepts are tracking abilities allowing the accumulation of knowledge about individuals, see (Millikan, 2000).

tend some particular ICS either via linking simple proposition-like structures (such as in (1), (3c), (7c), (11c), (14)), or, more locally, by attaching linked elaborations of nodes of any type (e.g. adjunct processing, see (11a)). At any point, either interlocutor can take over to realise the currently predicted goals in the ICS. This can be illustrated in the sharing of the dependency constrained by the locality definitive of reflexive anaphors:

(19) Mary: Did you burn Bob: myself? No.

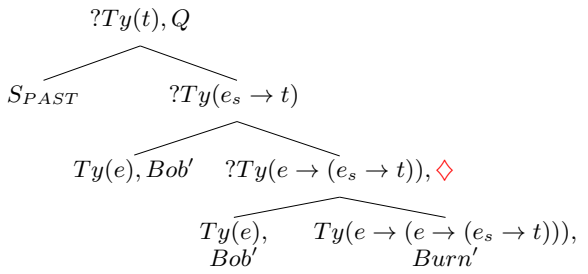
As shown in (19), Mary starts a query involving an indexical metavariable contributed by *you* that is resolved by reference to the $Hearer'$ contextual parameter currently occupied by Bob':

(20) $Mary$:*Did you burn* $\longmapsto$



With the ICS tracking the speaker/hearer roles as they shift subsententially, 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:

(21) Bob:*myself?* $\longmapsto$



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 according to any other syntactic/semantic framework.This successful result relies on (a) the lack of a syntactic level of representation, and (b) the subsentential licensing of contextual dependencies. In combination, these design features render the fact that the utterance constitutes a joint action irrelevant for the well-formedness of the sequence of actions constituting the string production.

This means that 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 a speaker-hearer temporary agentive unit performs in synchrony (Gregoromichelaki et al., 2011; Gregoromichelaki, 2013; Gregoromichelaki and Kempson, 2016) within a space of joint affordances.Given that parsing/production are joint predictive activities, driven by the participants' joint possible affordances, a current goal choice point in the DAG may be satisfied by a current hearer, so that it yields the retrieval/provision of conceptual information that matches satisfactorily the original speaker's needs or preferences, as in (7), (5), deflects the original speaker's action, (4), or can be judged to require some adjustment via backtracking that can be seamlessly and immediately provided by feedback extending/modifying the ensuing ICS, (3e), (15).

5 Conclusion

The dynamic articulation of DS, and its emphasis on incrementality and domain-generality of the processing mechanisms, reflect the formalism's intended cross-modal applicability in modelling uniformly NL grammars, action, and perception via a constitutive property of action: goal-directed predictivity. In our view, this commitment allows for parsimonious explanations of NL data and accommodates now standard psycholinguistic evidence of prediction from sentence processing studies (Altmann and Kamide, 1999; Trueswell and Tanenhaus, 2005, a.o.) as well as experimental data from multimodal, situated dialogue where notions of know-how, agent coupling, joint purpose, and direct perception replace the need for individualistic propositional-inferential theories (Mills and Gregoromichelaki, 2010; Shockley et al., 2009, a.o.). Gricean theories of common ground have placed a heavy burden on mindreading capacities as they separate syntactic and semantic knowledge from ac-

tion and perception. DS processing in contrast is able to take advantage of the temporally extended nature of processing at various scales because it assumes that NL know-how and practice-conforming behaviour can be uniformly modelled as meshing constraints without the necessary mediation of processing/inferring sentential/propositional units. Accordingly, there is no notion of wellformedness defined over sentence-proposition mappings, only systematicity/productivity grounded via the incremental, interaction-oriented NL procedures. Intra-individual NL mechanisms are incomplete on their own and need to be directed and constrained by affordances available in the sociomaterial environment. This complementarity ensures that NL elements acquire normative properties and effects contributing in turn to the establishment of novel practices that interleave seamlessly perceptual experiences, physical actions, and multimodal sources of information.

Acknowledgements

We want to thank our friends and collaborators in this project: Ellen Breitholtz, Ronnie Cann, Stergios Chatzikyriakidis, Robin Cooper, Arash Eshghi, Pat Healey, Julian Hough, Staffan Larsson, Gregory Mills, Matt Purver, Peter Sutton, and Graham White. We also owe thanks to the three anonymous reviewers for critical feedback on the manuscript that helped us clarify and develop our position more succinctly. Howes was supported by a grant from the Swedish Research Council (VR project 2014-39) for the establishment of the Centre for Linguistic Theory and Studies in Probability (CLASP) at the University of Gothenburg. All are gratefully acknowledged.

References

G. Altmann and Y. Kamide. Incremental interpretation at verbs: Restricting the domain of subsequent reference. *Cognition*, 73(3):247–264, 1999.

M. L. Anderson. *After Phrenology*. Cambridge University Press, Cambridge, 2014.

J. Barwise and J. Perry. *Situations and Attitudes*. MIT Press, Cambridge, MA, 1983.

J. B. Bavelas, L. Coates, and T. Johnson. Listeners as co-narrators. *Journal of Personality and Social Psychology*, 79(6):941–952, 2000.

M. H. Bickhard and D. M. Richie. *On The Nature Of Representation: A Case Study Of James Gibson's Theory Of Perception*. Praeger, New York, 1983.

J. Bruineberg, A. Chemero, and E. Rietveld. General ecological information supports engagement with affordances for 'higher' cognition. *Synthese*, 196 (12):5231–5251, 2018.

T. Buhrmann, E. A. Di Paolo, and X. Barandiaran. A Dynamical Systems Account of Sensorimotor Contingencies. *Frontiers in Psychology*, 4, May 2013. ISSN 1664-1078.

R. Cann, R. Kempson, and L. Marten. *The Dynamics of Language*. Elsevier, Oxford, 2005.

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

D. Casasanto and G. Lupyan. All concepts are ad hoc concepts. In *The conceptual mind: New directions in the study of concepts*, pages 543–566. MIT Press, Cambridge, MA, 2015.

A. Chemero. *Radical Embodied Cognitive Science*. MIT Press, Cambridge, MA, 2009.

P. Cisek and J. F. Kalaska. Neural Mechanisms for Interacting with a World Full of Action Choices. *Annual Review of Neuroscience*, 33(1):269–298, 2010.

A. Clark. Busting out: Predictive brains, embodied minds, and the puzzle of the evidentiary veil. *Noûs*, 51(4):727–753, 2017a.

A. Clark. How to knit your own markov blanket. In T. Wiese and W. Metzinger, editors, *Philosophy and Predictive Processing: 3. Frankfurt am Main: MIND Group.* Johannes Gutenberg-Universität Mainz, Frankfurt am Main, 2017b.

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

H. H. Clark. Wordless questions, wordless answers. In De Ruiter, Jan P, editor, *Questions: Formal, Functional and Interactional Perspectives*, pages 81–100. Cambridge University Press, Cambridge, 2012.

R. Cooper. Type theory and semantics in flux. In R. Kempson, N. Asher, and T. Fernando, editors, *Handbook of the Philosophy of Science*, volume 14, pages 271–323. Elsevier, 2012.

M. DeLanda. *Intensive Science and Virtual Philosophy*. Bloomsbury, London, 2013.

H. L. Dreyfus. The myth of the pervasiveness of the mental. In J. K. Schear, editor, *Mind, Reason, and Being-in-the-World*. Routledge, London, 2013.

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

A. Eshghi and O. Lemon. How domain-general can we be? learning incremental dialogue systems without dialogue acts. In *Proceedings of the 18th Workshop on the Semantics and Pragmatics of Dialogue*, pages 53–61, 2014.

A. Eshghi, M. Purver, and J. Hough. DyLan: Parser For Dynamic Syntax. Technical report, Queen Mary University of London, 2011.

A. Eshghi, M. Purver, and J. Hough. Probabilistic induction for an incremental semantic grammar. In *Proceedings of the 10th International Conference on Computational Semantics (IWCS)*, pages 107–118. ACL, 2013.

A. Eshghi, C. Howes, E. Gregoromichelaki, J. Hough, and M. Purver. Feedback in conversation as incremental semantic update. In *Proceedings of the 11th International Conference on Computational Semantics (IWCS)*, pages 261–271. ACL, 2015.

A. Eshghi, I. Shalyminov, and O. Lemon. Interactional dynamics and the emergence of language games. In C. Howes and H. Rieser, editors, *Proceedings of the workshop on Formal Approaches to the Dynamics of Linguistic Interaction (FADLI)*, number 1863 in CEUR Workshop Proceedings, Aachen, 2017.

N. H. Frijda, K. R. Ridderinkhof, and E. Rietveld. Impulsive action: Emotional impulses and their control. *Frontiers in Psychology*, 5, 2014.

A. Gargett, E. Gregoromichelaki, C. Howes, and Y. Sato. Dialogue-grammar correspondence in dynamic syntax. In *Proceedings of the 12th Workshop on the Semantics and Pragmatics of Dialogue*, 2008.

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

J. J. Gibson. *The Ecological Approach to Visual Perception: Classic Edition*. Psychology Press, New York, 2014.

J. Ginzburg. *The Interactive Stance: Meaning for Conversation*. Oxford University Press, Oxford, 2012.

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

E. Gregoromichelaki. Grammar as action in language and music. In M. Orwin, R. Kempson, and C. Howes, editors, *Language, Music and Interaction*, pages 93–134. College Publications, London, 2013.

E. Gregoromichelaki. Quotation in dialogue. In P. Saka and M. Johnson, editors, *The Semantics and Pragmatics of Quotation*, pages 195–255. Springer, Cham, 2017.

E. Gregoromichelaki and R. Kempson. Joint utterances and the (split-) turn taking puzzle. In J. L. Mey and A. Capone, editors, *Interdisciplinary studies in Pragmatics, Culture and Society*. Springer, Heidelberg, 2016.

E. Gregoromichelaki and R. Kempson. Procedural syntax. In R. Carston, B. Clark, and K. Scott, editors, *Relevance*. Cambridge Uuniversity Press, Cambridge, 2018.

E. Gregoromichelaki, Y. Sato, R. Kempson, G. A., and C. Howes. Dialogue modelling and the remit of core grammar. In *Proceedings of the 8th International Conference on Computational Semantics (IWCS)*, 2009.

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

E. Gregoromichelaki, R. Cann, and R. Kempson. On coordination in dialogue: Subsentential talk and its implications. In L. Goldstein, editor, *On Brevity*. Oxford University Press, Oxford, 2013.

E. Gregoromichelaki, C. Howes, A. Eshghi, R. Kempson, J. Hough, M. Sadrzadeh, M. Purver, and G. Wijnholds. Normativity, meaning plasticity, and the significance of vector space semantics. In *Proceedings of the 23rd Workshop on the Semantics and Pragmatics of Dialogue*, 2019.

H. Heft. Affordances and the body: An intentional analysis of Gibson's ecological approach to visual perception. *Journal for the Theory of Social Behaviour*, 19(1):1–30, 1989.

J. Hough. *Modelling Incremental Self-Repair Processing in Dialogue. PhD Thesis*. Queen Mary, University of London, 2015.

J. Hough and M. Purver. Probabilistic type theory for incremental dialogue processing. In *Proceedings of the EACL 2014 Workshop on Type Theory and Natural Language Semantics (TTNLS)*, pages 80–88. ACL, 2014.

C. Howes and A. Eshghi. Feedback relevance spaces: The organisation of increments in conversation. In *Proceedings of the 12th International Conference on Computational Semantics (IWCS) — Short papers*. ACL, 2017.

R. Kempson, W. Meyer-Viol, and D. Gabbay. *Dynamic Syntax*. Blackwell, Oxford, 2001.

R. Kempson, E. Gregoromichelaki, and C. Howes(eds.). *The Dynamics of Lexical Interfaces*. Studies in Constraint Based Lexicalism. CSLI, Stanford, CA, 2011.

R. Kempson, R. Cann, E. Gregoromichelaki, and S. Chatzikyriakidis. Language as mechanisms for interaction. *Theoretical Linguistics*, 42(3-4):203–276, 2016.

R. Kempson, R. Cann, E. Gregoromichelaki, and S. Chatzikyriakidis. Action-Based Grammar. *Theoretical Linguistics*, 43(1-2):141–167, 2017.

S. Kirby. *Function Selection and Innateness: The Emergence of Language Universals*. Cambridge University Press, Cambridge, 1999.

R. W. Langacker. Syntactic reanalysis. In C. N. Li, editor, *Mechanisms of Syntactic Change*, volume 58. University of Texas Press, Austin, 1977.

G. H. Lerner. On the place of linguistic resources in the organization of talk-in-interaction: Grammar as action in prompting a speaker to elaborate. *Research on Language and Social Interaction*, 37(2): 151–184, 2004.

P. Linell. *The Written Language Bias in Linguistics: Its Nature, Origins and Transformations*. Routledge, London, 2005.

D. Marr. *Vision: A Computational Investigation into the Human Representation and Processing of Visual Information*. WH Freeman and Company, San Francisco, 1982.

A. Maye and A. K. Engel. A discrete computational model of sensorimotor contingencies for object perception and control of behavior. In *2011 IEEE International Conference on Robotics and Automation*, pages 3810–3815, 2011.

R. G. Millikan. *On clear and confused ideas: An essay about substance concepts*. Cambridge University Press, Cambridge, 2000.

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

G. Mills and E. Gregoromichelaki. Establishing coherence in dialogue: Sequentiality, intentions and negotiation. In *Proceedings of the 14th Workshop on the Semantics and Pragmatics of Dialogue*, pages 17–24, 2010.

G. J. Mills. Dialogue in joint activity: Complementarity, convergence and conventionalization. *New Ideas in Psychology*, 32:158–173, 2014.

A. Noë. *Varieties of presence*. Harvard University Press Cambridge, MA, 2012.

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

M. Purver, E. Gregoromichelaki, W. Meyer-Viol, and R. Cann. Splitting the 'I's and crossing the 'you's: Context, speech acts and grammar. In *Proceedings of the 14th Workshop on the Semantics and Pragmatics of Dialogue*, pages 43–50, 2010.

M. Purver, A. Eshghi, and J. Hough. Incremental semantic construction in a dialogue system. In *Proceedings of the 9th International Conference on Computational Semantics (IWCS)*, 2011.

E. Rietveld, D. Denys, and M. Van Westen. Ecological-Enactive Cognition as engaging with a field of relevant affordances. In *The Oxford Handbook of 4E Cognition*, page 41. Oxford University Press, Oxford, 2018.

H. Sacks. *Lectures on Conversation*. Blackwell, Oxford, 1992.

Y. Sato. Local ambiguity, search strategies and parsing in dynamic syntax. In E. Gregoromichelaki, R. Kempson, and C. Howes, editors, *The Dynamics of Lexical Interfaces*. CSLI Publications, Stanford, 2011.

E. A. Schegloff. *Sequence organization in interaction: A primer in Conversation Analysis I*. Cambridge University Press, Cambridge, 2007.

K. Shockley, D. C. Richardson, and R. Dale. Conversation and Coordinative Structures. *Topics in Cognitive Science*, 1(2):305–319, 2009.

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

J. C. Trueswell and M. K. Tanenhaus. *Approaches to studying world-situated language use: Bridging the language-as-product and language-as-action traditions*. MIT Press, Cambridge, MA, 2005.

L. Wittgenstein. *Philosophical Investigations*. Blackwell, Oxford, 1953.

L. Wittgenstein. *Philosophical Grammar*. University of California Press, Berkley, CA, 2005.