Language and Cognition as Distributed Process Interactions

Eleni Gregoromichelaki
University of Gothenburg
eleni.gregoromichelaki@gu.se

Arash Eshghi
Heriot-Watt University
a.eshghi@hw.ac.uk

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

Gregory J. Mills
University of Groningen
g.j.mills@rug.nl

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

Julian Hough
Queen Mary University of London
j.hough@qmul.ac.uk

Patrick G. T. Healey
Queen Mary University of London
p.healey@qmul.ac.uk

Matthew Purver
Queen Mary University of London
m.purver@qmul.ac.uk

Abstract

In this position paper we argue that a conception of linguistic competence and conversational abilities that would fulfil the aims of Artificial General Intelligence cannot remain characterised as a static system of patterns induced from disembodied textual data. Instead, it should be modelled as a continuous, active, and interactive learning process. This is in line with the metaphysical and cognitive assumptions of Interactivism regarding the fundamental status of processes, as well as distributed cognition perspectives which argue that language does not reside in individual minds, brains, or bodies but is "spread out", embedded, and distributed in the available multimodal interactions with the environment. We show the usefulness of the formalism of Dynamic Syntax with Type Theory with Records (DS-TTR) in modelling dialogue to this end.

1 Introduction

Until recently, internalistic and static accounts of cognition have been the mainstream position in cognitive science and philosophy. However, dynamic accounts are now on the rise (Noë, 2004; Bickhard, 2009; Seibt, 2018; Manzotti and Chella, 2018, a.o.) alongside a growing interest in process metaphysics, substantiating the intuitive phenomenal idea of a dynamic, ever-changing reality while further justification is provided by recent relational interpretations of quantum mechanics (Laudisa and Rovelli, 2021) and category-theoretic results in mathematics (e.g. the Yoneda lemma, see Bradley et al., 2021). Moving away from the computational theory of mind with brain-internal representations and computations, current theories also argue that body–world interactions is what should be taken to constitute cognition (see, e.g. Hutchins, 1995).

In contrast, the idea of human language knowledge as an *abstract* and *static* system still underpins much work in theoretical linguistics, as well as language model architectures underlying recent impressive advances in NLP and AI (such as BERT (Devlin et al.), GPT3 (Brown et al., 2020) and their multimodal analogues e.g. ViLBERT (Lu et al., 2019), LXMERT (Tan and Bansal, 2019), Imagen (Saharia, 2022), DALL-E 2 (Ramesh et al., 2022), Gato (Reed et al, 2022) a.o.). The same view has been taken in computational dialogue modelling across the board, which retains the idea of human language knowledge as an autonomous and static system state. This system reconstructs human thought and communication as underpinned by module-internal rules and representations of a grammar and a lexicon enriched by some theory of mind module to explain performance. It is then natural to suggest that the system can be learned from static, disembodied textual data, and used for various downstream tasks after suitable fine-tuning.

Models implemented under this perspective have achieved great success in tasks that depend on reproducing patterns of very limited interactions with their environment (e.g., predicting upcoming input text), which allows learning of patterns of relationships among words. However, most researchers now concede that we have reached a point of diminishing returns (Bender and Koller, 2020). The constant increase of scale in amounts of data, computational resources, and parameters that are now required for minimal progress is un-

Proceedings of the 26th Workshop on the Semantics and Pragmatics of Dialogue, August, 22-24, 2022, Dublin.

sustainable both environmentally and due to the complexity of "black box" 'foundational models' (Bommasani et al., 2021). This results in lack of trust and confidence by users and the public due to the inscrutability and unexpected behaviours of current systems (see, e.g., Molnar, 2022).

As an alternative, we argue that language that fulfils the aims of AI and full artificial general intelligence (AGI) cannot be characterised as a static system of patterns induced from data as the result of learning but remaining unchanged during moment-to-moment (incremental) interaction with the user. Instead, language needs to be itself characterised as a *continuous, active,* and *interactive learning process*. This means that constant change and adaptation is what sustains any stable organisation we might detect during snapshot observations. This is in line with distributed cognition perspectives which argue that language is a system *property* that does not reside in individual minds, brains or bodies but is "spread out", embedded and distributed in the available multimodal interactions with the sociomaterial environment (cf also the Vygotskian robotics perspective e.g. Mirolli and Parisi, 2011).

One particularly acute symptom of the misconception of language as a static, representational system is that progress in creating natural interactions in conversational AI (aka 'dialogue systems') has plateaued. End-users of such systems have expectations of naturalness, intelligence, flexibility, and robustness to error, regularly leading to disappointment and frustration (Moore, 2017; Clark et al., 2019; Chaves and Gerosa, 2021; Luger and Sellen, 2016; Fischer et al., 2019). Large-scale end-to-end neural architectures (e.g. Wolf et al., 2019) display impressive capacities in terms of producing fluent immediate responses, but do not adequately capture human capacities in learning *appropriately adaptive* incremental conversational behaviours. Often such systems neglect the overall coherence of a situated dialogue setting thus lacking consistency with respect to the longer history of the dialogue and its future prospects with respect to achieving some goal (see e.g. Li et al., 2020; Vinyals and Le, 2015; Shang et al., 2015; Sordoni et al., 2015). As a consequence, today's conversational AI systems do not possess the strategic and embodied skills to negotiate the ambiguity, vagueness, and nuances of human-human conversation, and thus cannot learn and adapt to new people, tasks, and situations.

In this respect, critics of deep learning and current AI constantly point out that what is missing from such models is some notion of "semantics" to be articulated independently from the level of "forms", which is what is supposedly captured by such models (see, e.g. Bender and Koller, 2020; Bender et al., 2021). However, this criticism is only valid if it is taken for granted that there is such an objectively defined separation, i.e., form vs meaning, and, moreover, that AI systems of whatever variety are all meant to operate independently as autonomous cognitive agents. Alternatively, from the perspective of seeing language as a constructivist sociocultural process, form and function do not have to be distinguished but both of them can be seen as human abstractions of the epiphenomenal effects of underlying processes. Process organisation is what constitutes 'form' but such organisations are inherently functional. Given that processes interact and self-organise with emergent results at various levels (Bickhard, 2021), the autonomy of AI and NLP systems does not have to be taken as an all-or-nothing issue but as gradations of autonomy and independence depending on the purposes of use and the abilities of the agent. Unlike Piagetian constructivist views of human development, which arguably resemble the construals of current foundational models' learning regimes, Vygotskian cognitive robotics approaches to higher-level cognitive skills emphasise the 'internalisation' of social processes within individual minds transforming interpersonal processes to intrapersonal operations (e.g. Mirolli and Parisi, 2011, cf. Bruineberg and Rietveld 2019). This approach retains the primacy of the organism's interaction with the sociomaterial environment as the unifying factor of the relevant processual (self-)organisation while also accounting for autonomous performance. From this perspective, a language model that is taken to solipsistically receive and process inputs similarly to an isolated "brain-in-a-vat" does not provide an adequate basis for expecting human-level performance. However, text-to-image systems like DALL-E 2 and Imagen or generalist systems like GATO (Reed et al, 2022) that connect language with another modality like vision and operate across various tasks are a first demonstration that convincing linguistic performance is not due to an autonomous knowledge system performing "linguistic" tasks in isolation. Instead, the processes that constitute the *linguistic* organisation of a system, whether human-

human, or 'human-in-the-loop', comprise a mode of perception/action that structures the phenomenal world for other modalities deriving the social co-constructive nature of cognition. Thus moving towards more realistically embedded language models, implemented through artificial agents that interact more and more autonomously but under the normative forces imposed by the sociomaterial environment, sustains the possibility of eventually developing artificial general intelligence (AGI).

In this position paper, we set out the challenge of language as process (Gregoromichelaki, 2018; Gregoromichelaki et al., 2019, 2020b,a), rejecting the separation between form and meaning, syntax and semantics/pragmatics, or structure and function. We then reflect on the effects of incorporating the process of establishing coordination in social interactions into the core of the model itself.

2 The inadequacy of code models and Gricean mechanisms

Human communication is often characterised under the 'code model', namely, as one agent encoding and transmitting a message (the 'sender') to be decoded by another agent (the 'receiver'). This is an instance of the 'encodingism problem' in cognitive science as identified by (Bickhard, 2009) a.o. Successful communication is characterised as the hearer correctly discovering some preformed message which the speaker intended to convey. This basic assumption underlies most psychological and pragmatic theories of interaction including the Interactive Alignment Model (Pickering and Garrod, 2004), Gricean pragmatics and Relevance Theory (Sperber and Wilson, 1995) which assume an underlying literal meaning enhanced by context-specific pragmatic inferences to uncover the speaker's intention. But this approach has failed spectacularly to account for the complexity and subtlety of sense-making in human interaction (see e.g., Rączaszek-Leonardi et al., 2014; Fowler and Hodges, 2016).

This failure is because the actions of participants in dialogue form a system of coupled components (see e.g., De Jaegher and Di Paolo, 2007) so that *feedback mechanisms*, like constant error indication and adjustment, are crucial for the stability, maintenance, and self-organisation of the system. Given the moment-by-moment possibility and precariousness of action coordination, participants do not need explicit representations of their own or others' mental states, and nor do they need to converge on a shared 'code' or criteria of success. Instead, their conceptions and contributions need to be complementary to sustain a social practice whose normative character is defined externally to their own private or explicit rationalisations of their behaviour.

Rethinking our conception of successful communication away from shared codes puts the flexibility and dynamism of natural language (NL) at the heart of communication. As Healey et al. (2018b) state "[i]nstead of thinking of effective communication as formulating a "perfect" message, it becomes about finding optimal ways to uncover and address misunderstandings" (see also Healey et al., 2018a). We go further, and do not characterise these practices as uncovering 'misunderstanding' or 'miscommunication', terms which suggest that they are in opposition to some common understanding or common ground. Instead, we characterise successful coordination (i.e. system self-organisation, rather than "communication") as the local, incremental accommodation of inevitable and necessary perturbations in the emergent formation of a complex dynamical system enabling people's contributions to larger social organisations that constitute their ecological niche ('form of life').

From a psychological perspective, the rapidity and high incrementality of turn-taking exchanges in dialogue (Levinson and Torreira, 2015; Sacks et al., 1974) shows that intractable exhaustive reasoning about some optimal local outcome is not what participants aim for (cf. Frank and Goodman, 2012). Instead, practices of navigating through, and local adjustment to, an incrementally evolving landscape of *affordances* (Rietveld et al., 2018) provided by the ecological niche and participants' own actions, enable the forms of distributed cognition observed in dialogue (e.g. Dingemanse, 2020).

Transferring this insight to the domain of language technology, this assumption partially explains the limited success of language models in mimicking many aspects of human performance in dialogue, especially when it comes to coordination and adaptation. We attribute the substantial current shortcomings of such models to the limited variety of data they are exposed to, lack of the ability to actively interact with the data (cf. Li et al., 2017; Lewis et al., 2017), lack of feedback, lack of physical embodiment (see e.g. Pustejovsky and Krishnaswamy, 2021), and lack of a system of values (normativity) engendered through some moral framework (Hodges, 2022). We suggest that progress in modelling human dialogue and conversational AI requires a radical reconception of NLs as *mechanisms for (inter)action*.

Affordances and repair Under our interpretation, affordances are publicly available resources which trigger motivations for action within agents (*solicitations*, e.g. Dreyfus, 2013). Affordances are not, as standard, simply properties of the environment or agent-internal mechanisms (cf. Bickhard, 2009). Rather, they are relations between agent abilities and what the current sociomaterial environment 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. Interlocutors thus acquire a joint perspective as long as they operate as a system with autonomous self-organisation underpinned by prediction error minimisation (as modelled within the Free Energy Principle framework in its ecological/enactive interpretation, e.g., Bruineberg et al., 2018; Kiverstein et al., 2022). The local and shifting landscape of affordances and the state and abilities of the agents involved determine at each moment a demarcated 'field of affordances', i.e., a subset of the landscape of affordances that are perceived as relevant by the agents. This 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 macro-level. It also removes the need to define propositional structure substitutes to account for partial 'situation convention' transformations (Bickhard, 1980, forthcoming). Additionally, rather than modelling repair of intention recognition failures as phenomena in (1) and (2) are standardly characterised, this externalist and distributed perspective aims at modelling the strategically introduced public intention co-construction through the affordances of so-called 'repair mechanisms' (see also Haugh, 2008; Haugh and Obana, 2015; Arundale, 1999):

(1) (a) A: so …umm this afternoon …
(b) B: let's go watch a film
(c) A: yeah

(2) (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 KS1 789-791]

3 Form, meaning, and interaction

Looking at single individuals out of context, there are unlimited degrees of freedom available for realising action opportunities, which leads to intractability, especially in Gricean models where coordination is modelled as recursive mindreading. This limitation can be overcome by conceptualising conversational interaction as process organisation into a coherent system: when agents become coupled and subsumed under an emergent sociocognitive system, degrees of freedom are severely restricted due to the top-down constraints exercised on individuals to perform their particular role in the achievement of joint action (e.g. Deacon, 2011). This helps to locally constrain individual choices, without individuals having to necessarily conceptualise such choices or build matching models of reality inside their own heads (i.e. with the world taken to be its own "best model", (e.g. Brooks, 1990; Hutchins, 1995).

Mismatches in skills and information are necessary ingredients of such an emergent process of coordination and complementarity in action. While compatibilities between participants act as a channel for smooth, automatic navigation of aspects of a shared space of action opportunities (affordances), they also form the background for revealing divergences. These divergences constitute sources of scaffolded learning and thus require attention and work to sustain the interaction. The prerequisites and presuppositions of the interaction thus become "present-at-hand" (Heidegger in (Dreyfus, 1990)) and constitute sources of learning and development by "educating the attention" (Gibson, 1966) of agents allowing them to differentiate novel opportunities or threats in their joint environment. Divergences trigger 'solution probing' processes, where the interlocutors attempt to reorient the trajectory of the joint action towards its incrementally emerging joint goals. At these points, aspects of the interaction regarding what is "appropriate" in that particular sociocultural practice (social normativities) become available as experiences and training for the individual participants who are in this way enabled to learn and develop their skills through interactions scaffolded by the relevant practices and other agents' abilities and guidance (see, e.g. Steffensen et al., 2016).

Data from human-human dialogues, such as (3), provides evidence that participants can fluently interact, with emergent coordination, despite the fact that conversational exchanges are superficially full of "fragments", non-linguistic signs, disfluencies, and non-verbal signals such as gestures and gaze:

(3)
1. **J:** Can you think of any catalysts?
2. **A:** Er is it potassium permanganate?
3. **J:** <unclear>
4. **A:** What
5. **J:** Pla <pause> a duck billed
6. **A:** Pardon?
7. **J:** A duck billed
8. **A:** Platypus.
9. **J:** And it's not platypus it's <pause> sounds like a type of pen.
10. **A:** Platinum.
11. **J:** Right, platinum. [BNC; FMR 728-737]

As seen here, units of meaning are co-created incrementally (Gregoromichelaki et al., 2013; Kempson et al., 2016) by multiple interlocutors using incomplete utterances (e.g. line 7 – Purver et al., 2011), with phenomena such as cross-person compound contributions (where one person continues another's utterance, as in lines 7 and 8 – Lerner, 1991; Howes, 2012), repairs (e.g. the clarification requests in lines 4 and 6 – Sacks et al., 1974; Purver, 2004), and disfluencies (e.g. the pause and restart in line 9 – Hough, 2015) – seen as 'performance errors' in traditional linguistics – crucial in the co-construction of meaning.

In (3), a chemistry tutor (J) prompts a student (A) to answer the question in line 1, illustrating the divergence and convergence complementarity that is key to driving dialogue forwards. The social roles of teacher and student constrain the way in which their several responses are interpreted and this interplay and meshing of factors belies distinctions such as form vs meaning, communication vs thought or speaker vs listener. From a standard individualistic perspective, one can characterise the exchange as indicating that from J's perspective, A's response in line 2 diverges from the expected answer. A finally produces the expected answer (thus demonstrating convergence with J's expectations) in line 10. This is a valid way of describing the process and could be how a single participant or observer might rationalise or abstract the dialogue process into a narrative that they construct post hoc. This meta-perspective is arguably the one that prevailed in the construction of dialogue systems (e.g. Kopp and Krämer, 2021) in the era before end-to-end statistical models.

However, this view neglects the fact that both participants operate in a context (a 'teaching context') that imposes normative constraints in what their actions should be as they perform the roles assigned to them by the sociocultural convention: there are no 'teacher' or 'student' roles outside this socially-afforded context. This is not necessarily a conceptualisation that is explicit in any individuals' real-time consciousness but it is an effect of the 'habitus' (a set of embodied dispositions, *solicitations*, e.g. (Dreyfus, 2013), or *effectivities* (Turvey, 1992)) that agents have acquired through enculturation. The characterisation of the interactive potential here is similar to Bickhard's 'situation convention' with the difference that it is not grounded exclusively through the participants' internal understanding or awareness. The practice is enabled outside the agents' brain processes to constitutively include extended temporal, material, and spatial processes converging in the interaction. In its turn, the process organisation that constitutes the practice constitutes the participants' (temporary) identities and the action possibilities afforded to them.

The exchange of information in the sense of 'semantic information' assumed in model-theoretic, denotational, or referential semantics is not the purpose of the interaction. Neither are Gricean or Neo-Gricean norms relevant in the sense of trying to figure out a speaker's communicative and informative intention. Instead, the task, or language game, here seems very similar to the elicitation tasks that current 'foundational' models are confronted with: sometimes they are required to complete a NL prompt given some additional context, or to produce an image by taking advantage of their experience with 'forms' of text and images that they have sifted over and compressed in their parameters and architecture (cf. Marcus, 2022). The functioning of these form-based results is then to be normatively determined within the overarching language game, which for foundational models is set by human users, thus minimising the agential properties of the models.

In the current case, the overarching goal is set by J and A's agency is minimised in the sense that A's responses are normatively judged as appropriate by J. From J's perspective, A's response in line 2 does not achieve the joint normative goal of the student-teacher context which A finally produces in line 10, namely, to enable A to respond appropriately when the situation requires retrieval of the type of elements that can be characterised as 'catalysts'. The naming word here (*catalyst*) has both linguistic and non-verbal affordances that are both targeted by the tuition. Inability to proceed is explicitly conveyed by A's clarification requests which act as signals for J to produce prompts probing A's knowledge of word *forms* to induce the answer. After a cue in line 5 fails to elicit the required convergence, J exploits the predictability induced by the compound noun phrase *duck-billed platypus* to get A to produce the first syllables of the answer to the original question. Of course, J's purpose is not to just entrench word form associations with the word *catalyst* in A. Instead, it is taken for granted that the signs (forms) constituting the words have action implications for the constitution of A as a capable agent with respect to chemistry. Form and meaning then, or 'natural meaning' and 'non-natural meaning', are not separate categories but abstractions that in reality stand for qualitatively similar and interrelated processes within organisations of networks of affordances (Bickhard forthcoming cf. Skyrms, 2010).

Both participants' actions are subsumed under the context-specific normative perspective that their actions be relevant to the elicitation of some particular answer to the question posed by J, with both operating as a coherent system performing complementary actions towards that goal and compensating for each other's failings to contribute appropriately. This management of the divergent and convergent contexts is incrementally and locally managed, with a hierarchy of joint goals and subgoals emerging opportunistically. J and A can only have probabilistic expectations as to what they are required to do moment-by-moment and have to correct and adjust their performance based on the feedback received.

In this dialogue, there is an asymmetry between the speakers, as J is both the expert, and more powerful than A. In fact, this asymmetry is endemic, diagnostic of not just all child/adult (Duveen and Psaltis, 2013; Kunert et al., 2011) or expert/non-expert exchanges (Lu et al., 2007; Pilnick and Dingwall, 2011), but all interactions. Differences in experiences, cultural background, individual physiology and social communities all contribute to differences in our language use, meaning that we never share the "same" language as anybody we nevertheless successfully interact with (Clark, 1998). This raises an important practical question: How can we communicate successfully when individual differences in language use are not the exception but the norm?

We believe that the answer to this question relies on reconceptualising NL grammars as modelling a set of skills for interaction relative to social practices (Gregoromichelaki et al., 2019, 2020b), in common with distributed language models (Cowley, 2009) and the dialogical perspective (Linell, 2009) but within a formally articulated architecture that lends itself to implementation. We now sketch such a model.

4 DS-TTR

DS-TTR (Purver et al., 2010, 2011; Hough, 2015) is a system that combines the dynamic logic (PDL) architecture of Dynamic Syntax (DS, see e.g., Kempson et al., 2001; Cann et al., 2005) with probabilistic versions of Type Theory with Records (TTR, Cooper, 2005, forthcoming). TTR types are interpreted in DS-TTR in dynamic terms as affordances (Gregoromichelaki et al., 2019, 2020b; Eshghi et al., 2022), that is, type names are triggers for sets of PDL actions, just as syntactic/semantic categories in DS are labels for tree-building actions. Actions are expressed as probabilistically licensed transition events among the states of a dynamic system – see Fig. 1 where outgoing edges/actions from each node form a learnable (Eshghi et al., 2013) probability distribution conditioned on the current state. DS-TTR is thus 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 unviablity in view of more competitive alternatives (see Fig. 1). Words, morphology, and syntax are, in this way, all modelled as indicators of opportunities for (inter-)action (Gregoromichelaki, 2018; Gregoromichelaki et al., 2019, 2020b,a). Participants' opportunities for action and their perspectives are modelled in a unified model of the whole system. Interactions are modelled as incrementally opening up a range of options so that selected alternatives can be pursued either successfully or unsuccessfully: even though a processing path might be initially highly favoured, it might nevertheless lead to an impasse so that processing is aborted and backtracking to an earlier state is required (Sato, 2011) due to the changing conditions downstream.

As Fig. 1 shows, edges correspond to DS actions; and nodes correspond to states defined by their predictive potential for further actions. However, one might also take a coarser-grained view of the DAG with edges corresponding to words (sequences of computational actions followed by

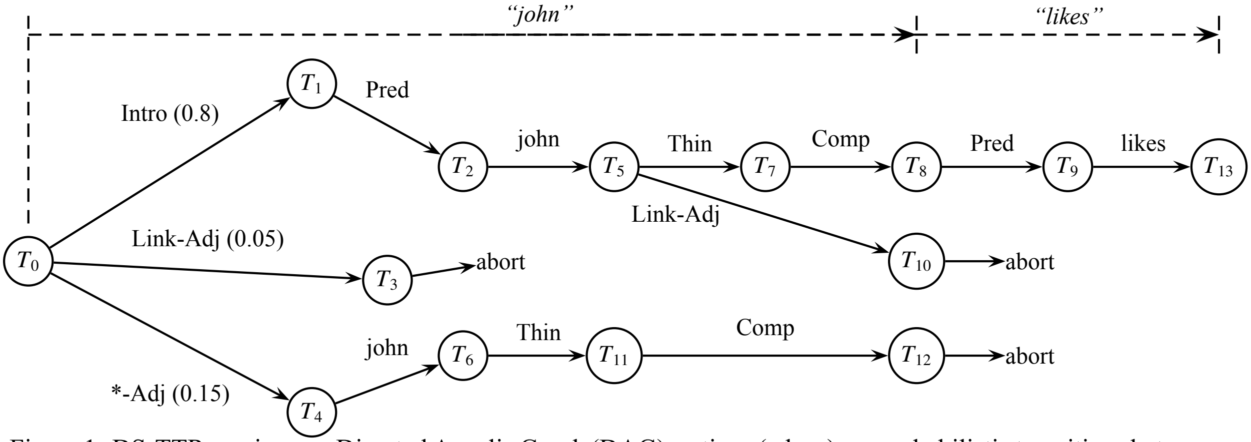

Figure 1: DS-TTR parsing as a Directed Acyclic Graph (DAG): actions (edges) are probabilistic transitions between partial trees (nodes).

a single lexical action) rather than single actions, and dropping abandoned parse paths (see Hough, 2015, for details).

On this view, DS-TTR parsing or generating a string of words or non-verbal tokens, induces some organisation of a state space of activity possibilities in combination with top-down actions ensuing from preexisting skills and dispositions of the participants involved (the 'grammar') (cf. Zadrozny, 2020). This either transforms the existing state space, adds new structural organisation to it, or removes existing paths through it. At each stage, a 'pointer' (◊) determines the local point of modification; and locally, the immediate path trajectory moves through a tree-shaped state space with nodes as states traversed by means of constraints expressed by the modal operators (e.g. $\langle\downarrow\rangle$, $\langle\uparrow\rangle$, $\langle\uparrow_*\rangle$) of a modal tree logic (the Logic of Finite Trees; LOFT: Blackburn and Meyer-Viol, 1994) expressing topological relations among current or future anticipated (i.e. predicted) nodes. The tree-shaped organisation of local processing trajectories reflects the conceptualisation structure induced by the unfolding utterance in terms of function-argument articulations. More globally, the state space is presented as a directed acyclic graph (DAG) that records possible paths of actions in a landscape defined by what the grammar, acting as a controller of the normativity pertaining to linguistic actions, allows as predictions of future interaction possibilities. The context required for processing various forms of context-dependency *is* the path searches provided by the DAG, augmented by affordances pertaining to the 'form of life' (e.g. Bruineberg et al., 2018) within which the interaction takes place.

Given the basic property of *predictivity* that sustains the DS-TTR mode of explanation of linguistic phenomena, the task confronting a DS-TTR learner is similar to the self-supervised language modelling task and even closer to current Reinforcement Learning (RL) architectures. Eshghi et al. (2017a,b) show how this idea can be implemented in narrow dialogue domains, where DS-TTR action policies are learned through exploring environmental contingencies (affordances) and acquiring skills in predicting suitable trajectories within the evolving landscape of affordances via RL methods. Hence, an induced DS-TTR grammar can be seen as a generative model capturing the interaction potential of a situational context, the latter including agents and sociomaterial constructs as in distributed cognition research.

5 Modelling feedback in DS-TTR

Given these inherent properties, DS-TTR has lent itself particularly well to dialogue modelling and analysis of dialogue phenomena within a unified architecture. Dialogue is modelled as the incremental and interactive composition of action sequences triggered by words either from oneself (in production) or an interlocutor (in comprehension) in an incrementally evolving context, the DAG past or future defined trajectories constituting the context, enabling unitary explanations of ellipsis (Kempson et al., 2015), self-repair (Hough and Purver, 2012; Hough, 2015), split utterances (Howes et al., 2011; Howes, 2012; Kempson et al., 2016), clarification requests (Gargett et al., 2009; Eshghi et al., 2015) and other feedback (Howes and Eshghi, 2021). In particular, it provides a basis for modelling backchannels (indications of agreement) vs clarification requests (overt indications of needing further development to enable agree-

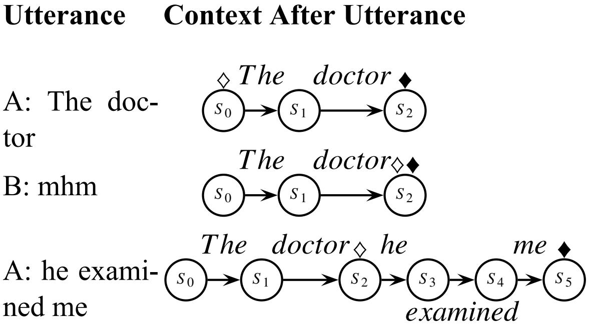

Figure 2: Backchannels as coordination pointers' movement in Interaction Control State-space (ICS)

ment), extensions vs corrections, hence 'repair', all as complementary procedural mechanisms for managing the types of transformations induced moment by moment in the ever evolving DAG space. As Eshghi et al. (2015) show, 'grounding' (the integration into the context of feedback) in a dyadic dialogue can be captured by including the perspective-relativisation of affordances: the DAG is augmented with two *coordination pointers*, the *self-pointer*, ♦, and the *other-pointer*, ◇, marking the points up to which the dialogue participants have each grounded the material. We dub this augmented context DAG, the *Interaction Control State-space* (ICS) - see Fig. 2.

Any utterance causes ICS pointer movement, and interlocutors each have their own ICS paths which can diverge, and re-converge as a result of clarification interaction and repair processes more generally. The self-pointer, ♦, on participant A's ICS tracks the point to which A has given evidence for reaching. The other-pointer, ◇, tracks where the other participant, B, has given evidence for reaching. For example, an utterance produced by A will move A's self-pointer on their own ICS to the rightmost node of their ICS; on B's ICS, it is the other-pointer that moves to the same location. On this model, the intersection of the path back to the ICS root from the self- and other-pointers is taken to be grounded, with the effect that parse or production search within this grounded pathway is precluded, thus removing the computational cost associated with finding alternative interpretation pathways, as well as formally explaining how conversations move forward.

This model has been shown to account for backchannels (Fig. 2), clarification interaction, and other-corrections (Eshghi et al., 2015; Howes and Eshghi, 2017, 2021). Clarification requests cause branching on the ICS, where the current path is abandoned and another branch constructed – a subsequent response plus the acknowledgement of this response eventually realigns the two coordination pointers, and the interlocutors' ICSs as a consequence (see Eshghi et al., 2015; Howes and Eshghi, 2021 for details). By contrast, backchannels and utterance continuations do not create new branches, but move the other-pointer forward on the current path.

6 DS-TTR in alignment with process and relational models of cognition and reality

DS-TTR and Interactivism (Bickhard, 2009, and elsewhere) share a lot in common. Both embrace the claim that the underlying foundation of linguistic theorising has to be reconsidered to a perspective that embraces the action-grounding and process metaphysics that standard representational frameworks have obscured. In this view, action dynamics are primary with processes being the most fundamental individuals (Seibt, 2018). Language processing is thus seen in both frameworks as transformation of a landscape of affordances (in DS-TTR terms) instead of decodings of denotational contents augmented by Gricean reasoning.

The two paradigms are thus strikingly congruent, yet they diverge in some respects. While agreeing that language is the fulcrum between what is mind-internal and -external, they diverge in the interpretation that they attribute to the process organisations that they invoke. The Bickhard view posits agent-internal representations (through *apperceptions*) as a necessary intermediate level of process in order to define error detectable by the agent. This is a crucial assumption for grounding Bickhard's notion of 'representation', albeit in non-standard dynamic terms. But from the DS-TTR point of view, this seems to presuppose that an agent has access only to its own dynamic mechanisms and processes, even when the agent is embedded in an overarching organisation like the one captured in a DS-TTR DAG. This means that the brain-internal perspective dominates the grounding of 'representation' even in such a ground-breaking dynamic model like the Interactivist model. In contrast, DS-TTR is more compatible with forms of radical realism, which construe the very existence of the objects of phenomenal experiences, including minds and languages, as products of interactions (e.g. Manzotti and Chella, 2018, cf. Laudisa and Rovelli, 2021;

Adlam and Rovelli, 2022), hence eliminating the need for a separate notion of mind-internal representations, without excluding them of course in certain circumstances. On this view, affordances are truly relational, generated and realised within distributed systems comprising multiple agents and within-agent levels. As in various forms of enactivism, social NL behaviours are understood as practices, with their normativity underpinned by a set of conditional actions (the 'grammar') inducing ongoing emergent flows that can be approximated, in more individualistic, abstract, and detached terms, as the often-studied notions of context, content, intentions, speech acts and the like. This radical extension of explanations of tools for use in communication as a core part of the grammar thus no longer corresponds to a capacity exclusively within the head of a single individual but is in some sense external to that, shared across participants. Moreover, the view of what an 'agent' is can be extended to non-biological artifacts, like artificial agents (Kockelman, 2011; Kiverstein et al., 2022). This is compatible with the view that process organisations are the fundamental explanatory factors of behaviours while metaphysical relationality implies that normativity can be attributed, albeit in a derivative sense, to the purposes of such agents (cf. Bickhard, 2021).

It is notable in this connection that the remit of data which DS-TTR is able and concerned to express corresponds remarkably closely to the insights of Conversational Analysis (CA), long widely ignored by theoretical linguists as doing no more than providing descriptions not amenable to formal characterisation, and in principle to be ignored due to merely constituting performance data (but cf. Ginzburg, 2012; Cooper, forthcoming).

Indeed the CA task was to provide a radically empiricist methodology to describe the interactions so characteristic of naturally occurring conversation. This can be given an internalist interpretation (cf. Ginzburg, 2012), but our aim here is to defuse the view that the skull or the human body provide a priori boundaries of where cognition, including grammars, is situated (cf. Albert and de Ruiter, 2018).

7 Future challenges

With grounding DS-TTR actions and types as affordances, there remains much work to be done, and at least one major problem. NLs universally display endemic context-dependence on the interpretations their words allow. Linguists are well aware of this fact, either addressing it by positing lexical ambiguities for every word of the language,[1] or attributing open-ended complexity of inference in the individual's capacity for language use. Against this challenge, the AI success in developing automated NL processing systems without any reference either to details of NL grammar formalisms or to such high-levels of inference stands in clear conflict with the abstract formalisms linguists have proposed – it is hard to envisage more damaging evidence against such approaches (Perconti and Plebe, 2019; Lappin, 2021). Much of this AI success has turned on large, neural language modelling techniques that instantiate the Firthian stance that the information-bearing load of words can be induced from the sets of words or affordances sharing the same local (multimodal) context window without any reference to intrinsic denotational content attributable to the words themselves (Gregoromichelaki et al., 2019).

In facing this challenge head on, work is currently exploring ways of combining the DS dynamic architecture with compositional Distributional Semantics tools (Purver et al., 2021). In this work, lexical items project *tensors* onto the interim emergent DS trees/states (instead of TTR record types), mapping onto vector spaces. This provides an explanatory basis from which the intrinsic non-determinism of lexical content can be modelled with content flexibility of NL expressions being essential to language variation and change (see, e.g., Gregoromichelaki et al., 2019). On this view, success in communication between participants is then predicted to rest in the emergent coordination due to the overlap shared by such spaces, for which feedback manifestly contributes as it conditions the shifting affordance landscape. This emergence, much in line with Bickhard's 'situation conventions' but externalised, plays a central role in refining emergent joint projects without requiring identity in understandings but, primarily, complementarity in action. Furthermore, work has been done in situating DS-TTR within embodied agents (Hough et al., 2020) giving non-verbal actions the same status as verbal utterances. Hence the claim that, far from defining a vehicle for communication leading to shared understanding of some defined denotational content, NL grammars are rather seen as comprising a set of skills for picking up interaction affordances within social practices.

[1] e.g., categorial grammar and its type polymorphism

References

Emily Adlam and Carlo Rovelli. 2022. Information is physical. *arXiv preprint arXiv:2203.13342*.

Saul Albert and J. P. de Ruiter. 2018. Repair: The interface between interaction and cognition. *Topics in Cognitive Science*, 10(2):279–313.

Robert B. Arundale. 1999. An alternative model and ideology of communication for an alternative to politeness theory. *Pragmatics*, 9(1):119–153.

Emily M. Bender, Timnit Gebru, Angelina McMillan-Major, and Shmargaret Shmitchell. 2021. On the Dangers of Stochastic Parrots. In *Proceedings of the 2021 ACM Conference*, pages 610–623, New York, NY, USA.

Emily M. Bender and Alexander Koller. 2020. Climbing towards NLU: On meaning, form, and understanding in the age of data. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 5185–5198, Online. Association for Computational Linguistics.

Mark H. Bickhard. 1980. *Cognition, Convention, and Communication*. Praeger.

Mark H Bickhard. 2009. The interactivist model. *Synthese*, 166(3):547–591.

Mark H. Bickhard. 2021. Emergent Mental Phenomena. In Robert W. Clowes, Klaus Gärtner, and Inês Hipólito, editors, *The Mind-Technology Problem*, pages 49–63. Springer, Cham.

Mark H. Bickhard. forthcoming. *The Whole Person*.

Patrick Blackburn and Wilfried Meyer-Viol. 1994. Linguistics, logic and finite trees. *Logic Journal of the Interest Group of Pure and Applied Logics*, 2(1):3–29.

Rishi Bommasani, Drew A. Hudson, and Ehsan Adeli et al. 2021. On the opportunities and risks of foundation models. *CoRR*, abs/2108.07258.

Tai-Danae Bradley, John Terilla, and Yiannis Vlassopoulos. 2021. An enriched category theory of language. *arXiv preprint arXiv:2106.07890*.

Rodney A Brooks. 1990. Elephants don't play chess. *Robotics and autonomous systems*, 6(1-2):3–15.

Tom B. Brown et al. 2020. Language models are few-shot learners. *arXiv:2205.06175*.

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

Jelle Bruineberg and Erik Rietveld. 2019. What's Inside Your Head Once You've Figured Out What Your Head's Inside Of. *Ecological Psychology*, 31(3):198–217.

Ronnie Cann, Ruth Kempson, and Lutz Marten. 2005. *The Dynamics of Language*. Elsevier, Oxford.

Ana Paula Chaves and Marco Aurelio Gerosa. 2021. How Should My Chatbot Interact. *International Journal of HCI*, 37(8):729–758.

Herbert H Clark. 1998. Communal lexicons. In Kirsten Malmkjær and John Williams, editors, *Context in Language Learning and Language Understanding*, chapter 4, pages 63–87. Cambridge University Press, Cambridge.

Leigh Clark, Nadia Pantidi, Orla Cooney, Philip Doyle, Diego Garaialde, Justin Edwards, Brendan Spillane, Emer Gilmartin, Christine Murad, Cosmin Munteanu, Vincent Wade, and Benjamin R. Cowan. 2019. What Makes a Good Conversation. In *Procs of the 2019 CHI Conference on Human Factors in Computing Systems*, pages 1–12. New York, NY, USA.

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

Robin Cooper. forthcoming. *From perception to communication*.

Stephen J Cowley. 2009. Distributed language and dynamics. *Pragmatics & Cognition*, 17(3):495–508.

Hanne De Jaegher and Ezequiel Di Paolo. 2007. Participatory sense-making. *Phenomenology and the Cognitive Sciences*, 6(4):485–507.

Terrence W Deacon. 2011. *Incomplete nature: How mind emerged from matter*. WW Norton & Company.

Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. BERT. In *Procs of NAACL*, pages 4171–4186.

Mark Dingemanse. 2020. Resource-rationality beyond individual minds. *BBS*, 43:e9.

Hubert L. Dreyfus. 1990. *Being-in-the-World*. MIT Press.

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

Gerard Duveen and Charis Psaltis. 2013. The constructive role of asymmetry in social interaction. In S. Moscovici, S. Jovchelovitch, and B. Wagoner, editors, *Development as a Social Processn*, pages 133–154. Routledge.

Arash Eshghi, Eleni Gregoromichelaki, and Christine Howes. 2022. Action Coordination and Learning in Dialogue. In *Probabilistic Approaches to Linguistic Theory*. CSLI.

Arash Eshghi, Christine Howes, Eleni Gregoromichelaki, Julian Hough, and Matt Purver. 2015. Feedback in conversation as incremental semantic update. In *Proceedings of the IWCS*, pages 261–271, London, UK. ACL.

Arash Eshghi, Matthew Purver, and Julian Hough. 2013. Probabilistic induction for an incremental semantic grammar. In *Proceedings of IWCS*, pages 107–118, Potsdam, Germany. ACL.

Arash Eshghi, Igor Shalyminov, and Oliver Lemon. 2017a. Bootstrapping incremental dialogue systems from minimal data: Linguistic knowledge or machine learning? In *Proceedings of EMNLP*, pages 2220–2230.

Arash Eshghi, Igor Shalyminov, and Oliver Lemon. 2017b. Interactional dynamics and the emergence of language games. In *CEUR Workshop Proceedings*, volume 1863, pages 17–21.

Joel E. Fischer, Stuart Reeves, Martin Porcheron, and Rein Ove Sikveland. 2019. Progressivity for voice interface design. In *CUI 19*.

Carol A. Fowler and Bert Hodges. 2016. Finding common ground. *New Ideas in Psychology*, 42:1–6.

Michael C. Frank and Noah D. Goodman. 2012. Predicting pragmatic reasoning in language games. *Science*, 336(6084):998–998.

Andrew Gargett, Eleni Gregoromichelaki, Ruth Kempson, Matthew Purver, and Yo Sato. 2009. Grammar resources for modelling dialogue dynamically. *Cognitive Neurodynamics*, 3(4):347–363.

James J. Gibson. 1966. *The Senses Considered as Perceptual Systems*. Houghton Mifflin.

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

Eleni Gregoromichelaki. 2018. Quotation in Dialogue. In Paul Saka and Michael Johnson, editors, *The Semantics and Pragmatics of Quotation*, pages 195–255. Springer.

Eleni Gregoromichelaki, Ronnie Cann, and Ruth Kempson. 2013. On coordination in dialogue. In Laurence Goldstein, editor, *On Brevity*. Oxford University Press, Oxford.

Eleni Gregoromichelaki, Stergios Chatzikyriakidis, Arash Eshghi, Julian Hough, Christine Howes, Ruth Kempson, Jieun Kiaer, Matthew Purver, Mehrnoosh Sadrzadeh, and Graham White. 2020a. Affordance competition in dialogue. In *Proceedings of the 24th SemDial*.

Eleni Gregoromichelaki, Christine Howes, Arash Eshghi, Ruth Kempson, Julian Hough, Mehrnoosh Sadrzadeh, Matthew Purver, and Gijs Wijnholds. 2019. Normativity, meaning plasticity, and the significance of Vector Space Semantics. In *Proceedings of the 23rd SemDial*.

Eleni Gregoromichelaki, Christine Howes, and Ruth Kempson. 2020b. Actionism in syntax and semantics. In *Dialogue and Perception DaP2018*, volume 2, pages 12–27.

Michael Haugh. 2008. Intention and diverging interpretings of implicature in the "uncovered meat" sermon. 5(2):201–228.

Michael Haugh and Yasuko Obana. 2015. Transformative continuations, (dis)affiliation, and accountability in Japanese interaction. *Text & Talk*, 35(5):597–619.

Patrick G. T. Healey, Gregory J. Mills, Arash Eshghi, and Christine Howes. 2018a. Running Repairs. *Topics in Cognitive Science*, 10(2):367–388.

Patrick G. T. Healey, Jan Peter de Ruiter, and Gregory J. Mills. 2018b. Editors' introduction: Miscommunication. *Topics in Cognitive Science*, 10(2).

Bert H. Hodges. 2022. Values Define Agency. *Adaptive Behavior*, page 10597123221076876.

Julian Hough. 2015. *Modelling Incremental Self-Repair Processing in Dialogue*. Ph.D. thesis, Queen Mary University of London.

Julian Hough, Lorenzo Jamone, David Schlangen, Guillaume Walck, and Robert Haschke. 2020. A types-as-classifiers approach to human-robot interaction for continuous structured state classification. *CLASP Papers in Computational Linguistics*, 2:28–40.

Julian Hough and Matthew Purver. 2012. Processing self-repairs in an incremental type-theoretic dialogue system. In *Proceedings of the 16th SemDial (SeineDial)*, pages 136–144.

Christine Howes. 2012. *Coordination in Dialogue: Using Compound Contributions to Join a Party*. Ph.D. thesis, Queen Mary University of London.

Christine Howes and Arash Eshghi. 2017. Feedback relevance spaces. In *IWCS 2017*.

Christine Howes and Arash Eshghi. 2021. Feedback Relevance Spaces: Interactional Constraints on Processing Contexts in Dynamic Syntax. *Journal of Logic, Language and Information*, 30(2):331–362.

Christine Howes, Matthew Purver, Patrick G. T. Healey, Gregory J. Mills, and Eleni Gregoromichelaki. 2011. On incrementality in dialogue: Evidence from compound contributions. *Dialogue and Discourse*, 2(1):279–311.

Edwin Hutchins. 1995. *Cognition in the Wild*. MIT Press.

Ruth Kempson, Ronnie Cann, Arash Eshghi, Eleni Gregoromichelaki, and Matthew Purver. 2015. Ellipsis. In Shalom Lappin and Chris Fox, editors, *The Handbook of Contemporary Semantic Theory*. Wiley-Blackwell.

Ruth Kempson, Ronnie Cann, Eleni Gregoromichelaki, and Stergios Chatzikiriakidis. 2016. Language as mechanisms for interaction. *Theoretical Linguistics*, 42(3-4):203–275.

Ruth Kempson, Wilfried Meyer-Viol, and Dov Gabbay. 2001. *Dynamic Syntax*. Blackwell.

Julian Kiverstein, Michael D. Kirchhoff, and Tom Froese. 2022. The Problem of Meaning: The Free Energy Principle and Artificial Agency. *Frontiers in Neurorobotics*, 16:844773.

Paul Kockelman. 2011. Biosemiosis, technocognition, and sociogenesisy. *Current Anthropology*, 52(5):711–739.

Stefan Kopp and Nicole Krämer. 2021. Revisiting Human-Agent Communication. *Frontiers in Psychology*, 12.

Richard Kunert, Raquel Fernández, and Willem Zuidema. 2011. Adaptation in child directed speech. In *Proc of SemDial (LosAngelogue)*, pages 112–119.

Shalom Lappin. 2021. *Deep Learning and Linguistic Representation*. Chapman and Hall.

Federico Laudisa and Carlo Rovelli. 2021. Relational Quantum Mechanics. In *The Stanford Encyclopedia of Philosophy*, Winter 2021 edition. Stanford University.

Gene H. Lerner. 1991. On the syntax of sentences-in-progress. *Language in Society*, pages 441–458.

Stephen C. Levinson and Francisco Torreira. 2015. Timing in turn-taking and its implications for processing models of language. *Frontiers in Psychology*, 6:731.

Mike Lewis, Denis Yarats, Yann Dauphin, Devi Parikh, and Dhruv Batra. 2017. Deal or No Deal? In *Proceedings of EMNLP*, pages 2443–2453.

Margaret Li, Stephen Roller, Ilia Kulikov, Sean Welleck, Y-Lan Boureau, Kyunghyun Cho, and Jason Weston. 2020. Don't Say That! In *Proceedings of the 58th ACL*, pages 4715–4728.

Xiujun Li, Yun-Nung Chen, Lihong Li, Jianfeng Gao, and Asli Celikyilmaz. 2017. End-to-end task-completion neural dialogue systems. In *Procs of the 8th NAACL*, pages 733–743.

Per Linell. 2009. *Rethinking Language, Mind, and World Dialogically*. IAP.

Jiasen Lu, Dhruv Batra, Devi Parikh, and Stefan Lee. 2019. Vilbert. In *Advances in Neural Information Processing Systems*, volume 32. Curran Associates, Inc.

Xin Lu, Barbara Di Eugenio, Trina C. Kershaw, Stellan Ohlsson, and Andrew Corrigan-Halpern. 2007. Expert vs. Non-expert Tutoring. In *Computational Linguistics and Intelligent Text Processing*, pages 456–467. Springer.

Ewa Luger and Abigail Sellen. 2016. "like having a really bad pa". In *Procs of the 2016 CHI Conference on Human Factors in Computing Systems*, CHI '16, page 5286–5297.

Riccardo Manzotti and Antonio Chella. 2018. Good old-fashioned artificial consciousness and the intermediate level fallacy. *Frontiers in Robotics and AI*, 5:39.

Gary Marcus. 2022. Horse rides astronaut. "`https://garymarcus.substack.com/p/horse-⊂rides-⊂astronaut?s=r`".

M. Mirolli and D. Parisi. 2011. Towards a Vygotskyan cognitive robotics. *New Ideas in Psychology*, 23(3):298–311.

Christoph Molnar. 2022. *Interpretable Machine Learning*, 2nd online edition.

Roger K. Moore. 2017. Is Spoken Language All-or-Nothing? In Kristiina Jokinen and Graham Wilcock, editors, *Dialogues with Social Robots*, pages 281–291. Springer.

Alva Noë. 2004. *Action in Perception*. MIT press.

Pietro Perconti and Alessio Plebe. 2019. Deep learning and embodiment. In *AIC*, pages 10–21.

Martin Pickering and Simon Garrod. 2004. Toward a mechanistic psychology of dialogue. *BBS*, 27:169–226.

Alison Pilnick and Robert Dingwall. 2011. On the remarkable persistence of asymmetry in doctor/patient interaction. *Social science & medicine*, 72(8):1374–1382.

Matthew Purver. 2004. *The Theory and Use of Clarification Requests in Dialogue*. Ph.D. thesis, University of London.

Matthew Purver, Arash Eshghi, and Julian Hough. 2011. Incremental semantic construction in a dialogue system. In *Proceedings of IWCS*, pages 365–369, Oxford, UK.

Matthew Purver, Eleni Gregoromichelaki, Wilfried Meyer-Viol, and Ronnie Cann. 2010. Splitting the 'I's and crossing the 'you's. In *Pros of the 14th SemDial*.

Matthew Purver, Mehrnoosh Sadrzadeh, Ruth Kempson, Gijs Wijnholds, and Julian Hough. 2021. Incremental composition in distributional semantics. *J. Log. Lang. Inf.*, 30(2):379–406.

James Pustejovsky and Nikhil Krishnaswamy. 2021. Embodied Human Computer Interaction. *KI - Künstliche Intelligenz*, 35(3):307–327.

Aditya Ramesh, Prafulla Dhariwal, Alex Nichol, Casey Chu, and Mark Chen. 2022. Hierarchical text-conditional image generation with clip latents. *arXiv preprint arXiv:2204.06125*.

Scott Reed et al. 2022. A generalist agent. *arXiv preprint arXiv:2205.06175*.

Erik Rietveld, Damiaan Denys, and Maarten Van Westen. 2018. Ecological-enactive cognition as engaging with a field of relevant affordances. In *The Oxford handbook of 4E cognition*, volume 41. OUP.

Joanna Rączaszek-Leonardi, Agnieszka Dębska, and Adam Sochanowicz. 2014. Pooling the ground. *Frontiers in Psychology*, 5:1233.

Harvey Sacks, Emmanuel A. Schegloff, and Gail Jefferson. 1974. A simplest systematics for the organization of turn-taking for conversation. *Language*, 50(4):696–735.

Chitwan et al Saharia. 2022. Photorealistic text-to-image diffusion models with deep language understanding. *arXiv preprint arXiv:2205.11487*.

Yo Sato. 2011. 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, CA.

Johanna Seibt. 2018. Ontological tools for the process turn in biology. In Nicholson and Dupré, editors, *Everything flows*, page 113. OUP.

Lifeng Shang, Zhengdong Lu, and Hang Li. 2015. Neural Responding Machine for Short-Text Conversation. *arXiv:1503.02364 [cs]*.

Brian Skyrms. 2010. *Signals*. OUP Oxford.

Alessandro Sordoni, Michel Galley, Michael Auli, Chris Brockett, Yangfeng Ji, Margaret Mitchell, Jian-Yun Nie, Jianfeng Gao, and Bill Dolan. 2015. A Neural Network Approach to Context-Sensitive Generation of Conversational Responses. *arXiv:1506.06714 [cs]*.

Dan Sperber and Deirdre Wilson. 1995. *Relevance*. Blackwell.

Sune Vork Steffensen, Frédéric Vallée-Tourangeau, and Gaëlle Vallée-Tourangeau. 2016. Cognitive events in a problem-solving task. *Journal of Cognitive Psychology*, 28(1):79–105.

Hao Tan and Mohit Bansal. 2019. LXMERT. In *Proceedings of EMNLP-IJCNLP*, pages 5100–5111.

Michael T Turvey. 1992. Affordances and prospective control. *Ecological psychology*, 4(3):173–187.

Oriol Vinyals and Quoc Le. 2015. A neural conversational model. *arXiv*, (1506.05869v3).

Thomas Wolf et al. 2019. TransferTransfo. *arXiv 1901.08149*.

Wlodek W. Zadrozny. 2020. Towards Coinductive Models for Natural Language Understanding. *arXiv:2012.05715*.