Metadata of the chapter that will be visualized in SpringerLink

Book Title	Language, Logic, and Computation
Series Title	
Chapter Title	Language as Mechanisms for Interaction: Towards an Evolutionary Tale
Copyright Year	2019
Copyright HolderName	Springer-Verlag GmbH Germany, part of Springer Nature

Author	Family Name	**Kempson**
	Particle	
	Given Name	**Ruth**
	Prefix	
	Suffix	
	Role	
	Division	
	Organization	King’s College London
	Address	London, UK
	Email	ruth.kempson@kcl.ac.uk
Corresponding Author	Family Name	**Gregoromichelaki**
	Particle	
	Given Name	**Eleni**
	Prefix	
	Suffix	
	Role	
	Division	
	Organization	King’s College London
	Address	London, UK
	Division	
	Organization	Heinrich Heine University
	Address	Düsseldorf, Germany
	Email	elenigregor@gmail.com
Author	Family Name	**Howes**
	Particle	
	Given Name	**Christine**
	Prefix	
	Suffix	
	Role	
	Division	
	Organization	University of Gothenburg
	Address	Gothenburg, Sweden
	Email	christine.howes@gu.se

Abstract	Scientific modelling of high-level cognitive processes has recently turned towards the implementation of philosophical/psychological views advocating enactive approaches to cognition.

Language as Mechanisms for Interaction: Towards an Evolutionary Tale

Ruth Kempson[1], Eleni Gregoromichelaki[1,2(✉)], and Christine Howes[3]

[1] King's College London, London, UK
ruth.kempson@kcl.ac.uk
[2] Heinrich Heine University, Düsseldorf, Germany
elenigregor@gmail.com
[3] University of Gothenburg, Gothenburg, Sweden
christine.howes@gu.se

1 The Isolationist Background

AQ1

Scientific modelling of high-level cognitive processes has recently turned towards the implementation of philosophical/psychological views advocating enactive approaches to cognition. Accordingly, for some time now, there has been a growing shift of emphasis in disciplines across cognitive science away from static representations of structure/content towards the dynamic, process-oriented modelling of skills and abilities that organisms employ, adjust, and perfect in dealing adequately with the ever-changing possibilities for action the environment affords [2,3,13,14,32,35,86]. In contrast, formal theorising about natural language (NL) has typically retained its characterisation as a code, an abstract system of rules and representations arbitrarily mapping forms to meanings. In this view, linguistic knowledge is codified as a 'grammar' mediating fixed mappings of phonological, syntactic and semantic representations. It is well-known that this characterisation is inadequate for any realistic application of this purported knowledge in a dynamically changing environment, issues pertaining to such applications being in principle precluded. As a result, such views are either presented as theories of an encapsulated cognitive capacity (as I-language or "competence": [10]) or are supplemented by invoking pragmatic competence underpinned by innate mechanisms of mindreading and altruism to bridge the gap [11,87]. Both solutions are undesirable if the aim is to account for basic NL properties. First, as we argue, the effects characterized as mindreading and altruism/cooperation are outcomes of the mechanisms that NLs instantiate, not themselves causes. Second, modelling of NLs as codes presupposes a synchronic and static view. This view has had a troubled ride in probing NL evolution. If a domain-specific, encapsulated capacity with arbitrary relations between levels of structure is assumed, gradualist accounts of NL evolution are precluded. Instead, the emergence of NL has been seen as a mutation, a so-called "sudden switch" (for a recent variant see [5]). In psychology, the code view precludes mechanisms that are subject to constant change and adaptation in response to events in the environment so that learning, plasticity, and cultural transmission are excluded AQ2

A. Silva et al. (Eds.): TbiLLC 2018, LNCS 11456, pp. 1–18, 2019.
https://doi.org/10.1007/978-3-662-59565-7_11

in favour of biological determination (nativism) and prespecified unfolding of capacities ("maturation"). Accordingly, learning one's native language has been seen as requiring a Language Acquisition Device in which the child hypothesises a succession of grammars, increasingly approximating the adult grammar.

2 Interaction and Natural Language

But even a model of the most basic and mundane uses of NL, i.e. communication via conversational interaction, is beyond the static, isolationist accounts. In psycholinguistics, the code model methodology enriched with mind-reading capacities presuppose what has been characterized as the "cognitive sandwich" view [53]. According to this view, the mind is structured at three levels: perception and action are seen as separate from each other and peripheral; cognition, the locus of propositional thought, planning, and executive control, stands in between as the filling. Applied to the modelling of dialogue, this view postulates that low-level perception and action involve a series of independent coding/decoding modules which are separate from and coordinated by the higher processes of cognition. Communication then is explicated as the transfer or sharing of "meanings", conceived as propositions, from one individual mind to another. As a result, the perceiver/listener is modelled not as an interacting agent/actor but as a passive recipient decoding stimuli produced by a speaker and replicating the speaker's thought. This view contradicts empirical observations of dialogue data showing that production and comprehension in dialogue are as tightly interwoven as argued in current computational neuroscience models linking action, action perception, and joint action [2,13]. The most glaring case is data showing rapid exchange of speaker/hearer roles in conversation even within the building of a single structure:

(1) A: We're going to
B: Marlborough
C: Marlborough?
B: to see Granny
C: With the dogs?
A: if you can keep them under control

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

(3) Jack: I just returned
Kathy: from
Jack: Finland [data from 63]

The existence of such interactive constructing of utterances/meanings is problematic for all conventional grammars, since any dependencies are able to be split apart so that resolution is only possible across the turn divide (as in (2) and (3)). These interactions also show how the direction they may take is by no means confined to realising some over-arching intentionally held content anticipated in advance of the speech event. And, of course, constructions such as these may be uttered by a single interlocutor, for their own benefit or their interlocutors', refining and elaborating an initial sentence/thought, as in (4):

(4) Mary's back. Late last night. From the US. Tired and frustrated.

Not being a cognitively demanding task, even language-acquiring infants can join in, adding to a proffered frame, (5), or initiating a frame construction process, (6)–(7):

(5) A: Old MacDonald had a farm, E-I-E-I-O. On that farm he had a
B: cow.

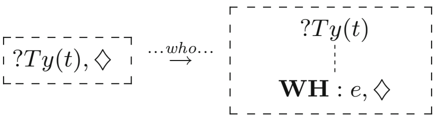
(6) *(2 year old on mum's bike waving at empty mooring over the canal)*
Eliot: Daddy!
Mother: That's right dear, you were here yesterday with Daddy clearing out the boat. [direct observation]

(7) A: Bear.
B: That's right dear, a panda.

The effect is one of rich potential for interactivity between participants, available from the earliest stages of NL development. Accordingly, in our view, what is needed to model such data is a grammar in which mechanisms of processing (actions) are modelled, rather than fixed mappings among form-contents. This involves a radical shift of assumptions, a shift that has been adopted in Dynamic Syntax – to which we now turn.

3 Language as Action

Dynamic Syntax (DS, [8,58]) is a grammar architecture whose core notion is incremental interpretation of word-sequences (comprehension) or linearisation of contents (production) relative to context. The DS syntactic engine, including the lexicon, is articulated in terms of goal-driven actions accomplished either by giving rise to expectations of further actions, by consuming contextual input, or by being abandoned as unviable in view of more competitive alternatives. Thus words, syntax, and morphology are all modelled as "affordances", opportunities for (inter-) action produced and recognised by interlocutors to perform step-by-step a coordinated mapping from perceivable stimuli (phonological strings) to conceptual mechanisms or vice-versa. To illustrate, we display below the (condensed) steps involved in the parsing of a standard long-distance dependency, *Who hugged Mary?*:[1]

(8) 

The task starts with a set of probabilistically-weighted predicted *inter-action-control states* (ICSs) represented as a directed acyclic graph (DAG) keeping

[1] The detailed justification of this formalism as a grammar formalism is given elsewhere ([8,9,37,56,58], and others).

track of how alternatives unfold and are progressively abandoned (we show only one snapshot of an active DAG path above and only the syntactically-relevant part). Such ICSs include salient environmental information, means of coordination, e.g. "repair" [21], and the recent history of processing. On this basis, they induce triggering of goals to build/linearise conceptual mechanisms ('ad-hoc concepts') classified as belonging to ontological types (e for entities in general, e_s for events, $e \rightarrow (e_s \rightarrow t)$ for predicates, etc). In (9) above, the goal is realized as a prediction to process next a proposition of type t. This is shown as a one-node tree with the prediction $?Ty(t)$ and the ICS's current focus of attention, the pointer $\Diamond$. Such initiating predictions can be of any type since the model aims to integrate predictions generated through any multimodal means with linguistic "fragments" (see e.g. (6)–(7) earlier) seamlessly induced within such DAG frames [43]. Additionally, no extra discourse levels or machinery like QUD [36] or DRSs is needed since the predictions are generated by the totality of information available to the DAG path state (cf. [29,61,65]). With the pointer at a node including a predicted outcome, predictions of further affordances/subgoals are generated under the expectation of eventual satisfaction of the current goal either by the processing of (verbal) input (as a hearer) or by producing that input (as a speaker). For (8), one of the probabilistically-licensed next steps for English (executed by defined lexical and general computational *macros* of actions) is displayed in the second partial tree therein: a prediction that a structurally underspecified node (indicated by the dotted line) can be built and can accommodate the result of parsing/generating *who*. As illustrated here, temporary uncertainty about the eventual contribution of some element is implemented through *structural underspecification*. Initially "unfixed" tree-nodes model the retention of the contribution of the *wh*-element in a memory buffer until it can unify with some argument node in the upcoming local domain. Non-referential words like *who* and other semantically underspecified elements (e.g. pronominals, anaphors, auxiliaries, tenses) contribute underspecified placeholders in the form of so-called *metavariables* (indicated in bold font). Metavariables in turn trigger search for their eventual type-compatible substitution from among contextually-salient entities or predicates.

General computational and lexically-triggered macros always intersperse to develop a conceptual graph of available affordances, locally taking the form of a binary tree: in (9), the verb contributes both conceptual structure in the form of unfolding the tree further, fetching an ad-hoc concept (indicated as Hug') developed according to contextual restrictions,[2] as well as placeholder metavariables for time and event entities to be supplied by the current ICS. Such conceptual structure is indefinitely extendible (see [18]) and "non-reconstructive" in the sense that it is not meant as an inner model of the world [43], (see also [14,15]). Instead, these structures function as 'interaction outcome indicators' [6] triggering possibilities of further (mental or physical) action, by either participant,

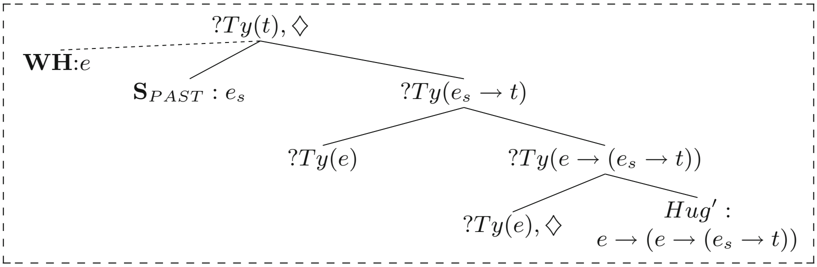
[2] In [22,25,27], this is modelled via a mapping onto a Type Theory with Records formulation, but we suppress these details here: see also [49,50,79].

extending or "repairing" the node elements, thus coordinating behaviour with selected aspects of the environment and each other.

Returning to the processing in (9), now NL-specific constraints kick in since the pointer $\Diamond$ is left at the argument node implementing the word-order restriction in English that the object needs to follow the verb:

(9)

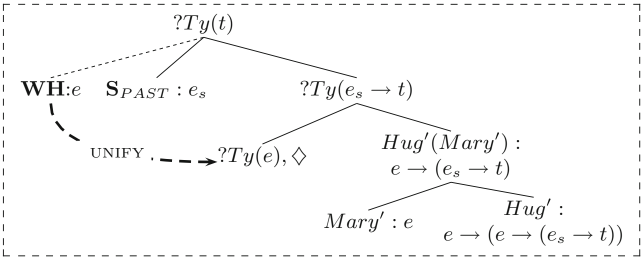
$\xrightarrow{...hugged...}$



At this point, the word *Mary* can be processed to initiate the tracking of a contextually-identifiable individual ($Mary'$) at the argument node internal to the predicate (for the view that such concepts are skill adaptations allowing the accumulation of knowledge about individuals, see [67]). After this step, everything is in place for the structural underspecification to be resolved, namely, the node annotated by *who* can now unify with the subject node of the predicate, which results in an ICS that includes the minimal content of an utterance of *Who hugged Mary?* imposed as a goal ($?Q_{\mathbf{WH}}$) for the next action steps (either by the speaker or the hearer):

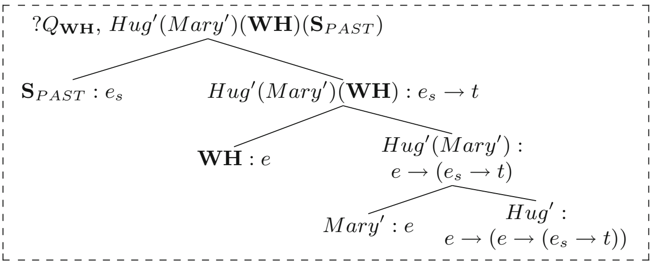
(10)

$\xrightarrow{...Mary...\texttt{unification macro}...}$



(11)

$\xrightarrow{...\texttt{tree-completion macros}...}$

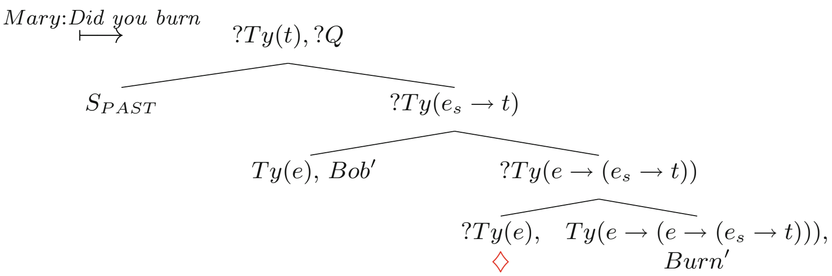


The DS model assumes tight interlinking of NL perception and action: the predictions generating the sequence of trees above are equally deployed in comprehension and production. *Comprehension* involves the generation of predictions/goals and awaiting input to satisfy them, while *production* involves the deployment of action (verbalising) by the predictor themselves in order to satisfy their predicted goals. By imposing top-down predictive and goal-directed processing at all comprehension/production stages, interlocutor feedback is constantly anticipated and seamlessly integrated in the ICS [21,33,34,39,76]. Given that the total information state is modelled by the DAG as a single level, even "semantically empty" feedback such as backchannels (*mmm*, *uh-huh*), serves, through the same mechanisms that implement lexical and computational actions, the function of continually realigning interlocutors' processing contexts, whilst ensuring that the problem of maintaining coordination is computationally tractable [22,23,44,51]. Integration in the ICS can involve adding simple proposition-like structures such as (11) or locally linked structures of any type incrementally elaborating some node of a tree in the ICS. For this reason, maintaining even abandoned options is required to achieve the concurrent modelling of conversational phenomena like clarification, self/other-corrections etc, but also, quotation, code-switching, humorous effects and puns [39,49]. Given the modelling of word-by-word incrementality, 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:

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

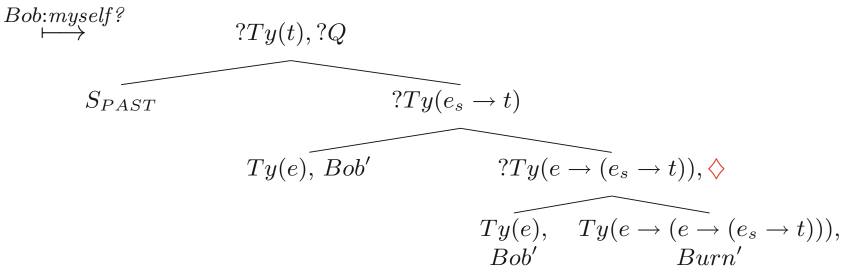
As shown in (12), Mary starts a query involving an indexical metavariable contributed by *you* that is resolved by reference to the *Hearer′* contextual parameter currently occupied by *Bob′*:

(13)



With the ICS tracking the speaker/hearer roles as they shift subsententially, these roles are reset in the next step when it happens that Bob takes over the utterance. *Myself* is then uttered. Being a pronominal, it contributes a metavariable and, being a reflexive indexical, it imposes the restriction that the entity to substitute that metavariable needs to be a co-argument that bears the *Speaker′* role. At this point in time, the only such available entity in context is again *Bob′* which is duly selected as the substituent of the metavariable:

(14)



As a result, binding of the reflexive is semantically appropriate, and locality is respected even though joining the string as a single sentence would be ungrammatical. This successful result relies on (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 wellformedness 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 speakers/hearers perform in synchrony [38,40,42]. As DS models syntax as a process (actions), not the resulting product, semantically equivalent strings might result in identical trees but the record of the processes on the DAG will be distinct. Syntactic alignment [74, a.o.] and priming in experimental data is explained by the reuse of these actions [57] while the fact that all conversational processing phenomena are modelled in a single level explains the finding that such alignment occurs at levels below chance in general conversation [45]. Due to subsentential step-by-step licensing, speakers are not required to plan propositional units, so hearers can perform "repair" subsententially without need to reason about propositional intentions. Given that parsing/production are predictive activities [61,75], a current goal in the ICS may be satisfied by a current hearer, so that it yields the retrieval/provision of conceptual information that matches satisfactorily the original speaker's goals, as in (2)–(5), or be judged to require some adjustment that can be seamlessly and immediately provided by feedback extending/modifying the ensuing ICS:

(15) Ken: He said all the colored people uh walk- walk down the street
and they may be all dressed up or somethin and these guys
eh white- white guys'll come by with .hh
Louise: mud.
Ken: mud, ink or anything and throw it at 'em [from 62]

The action dynamics in DS, and its emphasis on underspecification and update for both NL and context, reflect the formalism's fundamental cross-modal predictivity and integration of normative constraints from various sources, e.g. turn-taking conventions, within a single graph. This allows for parsimonious explanations of now standard psycholinguistic evidence of prediction from sentence processing studies [1,89, a.o.] without requiring internally structured predictive models (see [41] for comparison with [75]). The single-level assumption also allows for the fine-grained modelling of results of current turn-manipulation experiments showing that how people respond to truncated turns depends on how predictable the continuation is. Extremely predictable continuations do not even need to be articulated by either party in order to be taken as part of the interpretation, and continuations that are predictable in terms of structure but not content (such as those within a noun phrase) prompt dialogue participants to provide multi-functional utterances serving both as continuations and offering feedback as clarification requests ([52], cf. [29]). DS processing can model all these options since there is no notion of wellformedness defined over sentence-proposition mappings, only systematicity/productivity of procedures for incremental processing. Therefore, unlike non-incremental formalisms where explanation for these phenomena has to either be devolved to a parser external to the grammar or be relegated to performance "errors", fragmentary linguistic input/output and "repair" processes are not modelled as a problem for the interlocutors. Instead such processing is basic and constitutes the purpose of interaction which is to modify the interlocutors' cognitive and physical environments, a basic feature for learning and adaptation purposes.

4 Interaction and Language Learning

DS is a grammar modelling goal-directed coordination activity, with syntax transformed into a set of conditional update actions that induce or develop the processing environment of interacting agents. Depending on the moment-by-moment system-generated predictions of the next system state, processing strategies either attend to input verifying the predictions (as a hearer) or induce physical action realising the predictions (as a speaker), each agent manipulating the grammar mechanisms relative to their own capabilities, needs, desires, and goals but also as part of a system of coordinated processes with emergent properties [41,42]. This is unlike other frameworks [47,96] where these two apparently inverse activities are modelled as distinct, leaving mediating higher-order inference as the only means of modelling emergence and adaptation to the interlocutor's processing. DS, to the contrary, allows for variability without disruption in the affordances each agent perceives/pursues and, equally, divergence in what it is that they establish as the outcome of the interaction. Consequently, the DS update mechanisms have been shown to be learnable from child-directed semantically annotated data [25,26], where such asymmetries are crucial, and in the automatic induction of successful strategies serving task-specific dialogue games [24,27,54,81,94].

Given the embeddability of NL under domain-general skills and constraints as modelled by DS, learning an NL comes under one and the same domain of behavioural control, the establishment of sensorimotor contingencies, resulting from environmental and self-generated feedback [73,90]. In turn, given that the cultural/social environment is the main source of such contingencies for NL acquisition, the starting point for it is the now familiar observation (see [4]) that all utterance exchanges will necessarily involve moment by moment interaction between participants in communicative activities as they severally adjust jointly established action-control states to each other's desirable/undesirable affordances. Even from 4 months old, children enjoy interactive rituals like peekaboo games, in which there is no essential attribution of content, actively participating from 6.5 months (observation spring 2018). Moreover, [48] and [16] report on prelinguistic stages in which the caregiver characteristically uses their language fluently, invariably providing shaping feedback to the prelinguistic vocal behaviour of the child within their own conception of what this activity leads to (*mindshaping*: [95]). The result is an interactive effect between child and caregiver even in the absence of any expectation of mutual content duplication, the sole reward for both being the rich emotional bonding achieved by this interactive behaviour [88].

In the next phase of acquisition, the one-word utterance stage, successful communication again builds on the interaction between participants, despite, indeed riding upon, the asymmetry between them. This interactive behaviour rests on the reiteration of exchanges during which child and adult severally interpret what is offered them and engage in overt action to shared effect. As the child comes to isolate and so offer one-word utterances, there are notable structural patterns. On the one hand, the child may be offering some completion to a structured routine affordance just provided by the adult's utterance as in the nursery-rhyme exchange of (5). Or, given the coupling of the producer/comprehender systems, if the child is initiating some exchange with such a fragment, they can do so on the expectation that the adult will then develop it (as in (6) above). There are, furthermore, 'embedded correction' cases which can add elaboration, adjunct-like, to the child's offering, as in (7) [80]. In these earliest occurrences of NL, both producing and parsing such a construction would, following the anticipatory DS dynamic, involve predicting feedback and subsequently adjusting expectations (weightings), in the child's case, in favour of the carer's input stimulus. Learning an NL then is learning to exploit the affordances offered by interlocutors. In recurrent occurrences of such scenarios, sensorimotor contingencies will become entrenched so no matter how disjoint or asymmetric the construals of these events by the participants may be, at each stage, neither high-level inferences nor other-self mind-reading abductions need to be invoked (cf. [31,84]).

In fact, this asymmetry between participants in what they bring to bear in the conversational exchange continues across the lifespan, diagnostic of not just all expert/non-expert exchanges, but all dialogic encounters, as differences in experiences, cultural background, individual physiology and social communities all contribute to differences in people's language use, meaning that we never have the "same" language as anybody we are interacting with [17]. In consequence, variation and uncertainty lie at the heart of NL processing, and do not in general inhibit it. In any case, should such uncertainty be picked up on as problematic, NLs have tools specifically reifying the interaction and indicating need of clarification, correction, etc [19] so the pinpoint of uncertainty, if recognised as a hurdle, serves only to enrich the ongoing interaction by making explicit the implicit adjustment mechanisms of prediction-generation, which gradually becomes the basis for (explicit) inference and logical reasoning (instead of interaction relying on such capacities).

5 NL Evolution

The need for adjustment and change shapes all properties of NLs so that a dynamic perspective on NL abilities, rather than fixed form-content mappings, seems to us necessary. At all levels of interaction organisation, instead of high-level inferential processes deciphering hidden speaker intentions, it is domain-specific interaction patterns (*language games*) that allow for particular procedural conventions (i.e., in our view, words, syntax, semantics) to be modelled as emergent, learned, and adjusted during interaction [43,68,69]. This is possible because DS does not impose a single set of actions that must apply invariably to achieve each dialogue goal nor encoded speech act specifications modelling explicit propositional goals [40]. Instead, the composability of complex routines (*macros*) out of basic atomic actions can lead, through affordance competition [12], to plastic strategies that can be (re)deployed and refined at each instance to achieve results, with selection depending on the intersubjective processing environment structuring an interaction according to current needs. In this respect, computational work confirms the successful induction of domain-specific dialogue structures (language games) from very small amounts of unannotated data, with no dialogue act annotation but using instead a combination of DS and Reinforcement Learning [54,93]. Under such learning, multiple processing routes to the same dialogue goal are reinforced or inhibited by feedback depending on the situation and each individual's needs. Establishment of routinised processing also depends on efficiency in securing predictive success at minimal cost. For example, during reinforcement learning, the reward function upon reaching a goal penalises increasing dialogue time/length. This is a general constraint imposed by the organisation of the cognitive system itself, a property often invoked in linguistic pragmatics as determining communicative success [84], but here seen as the natural emergence of the way an organism has evolved to determine success in its task in manageable real time (the "lazy-brain hypothesis", [13]). This general view substitutes feedback-enhanced trial-and-error processing, selecting efficient routines through (inter)action, as the learning mechanism, in place of the need for internal world models and costly computation of others' internal mental states or common ground. Across multiple interactive situations this means that NL users can employ different strategies with different partners to reach the same outcome depending on their histories of interactions without local coordination failure. Long term, such tolerance of alternatives becomes a source of variation of the kind necessary for evolutionary selection.

Taking the fine details of such procedures to be the object of selection requires a view of evolution that is not confined to genetic modification [71,92]. Adjustments made possible by mechanisms loosely described as enculturation, niche construction, social learning, and cultural transmission [60] are involved here. Given that what is constructed during NL interactions are not world-mirroring models but repertoires of (inter)action-control states generating the next predicted inputs, DS follows the pattern assumed by enactive [72,86] and recent predictive coding models [14,15] for whole cognitive systems. In the latter, total brain-body organisms are described as instantiating predictive systems using previous experience at every step to anticipate with uncertainty the structure of the next incoming sensory array: perception, action (and imagination) all rely on probability distributions, rather than fixed decodings, over the incoming stimuli with different reliability weightings determining the ensuing adjustments as responses to error signals. Such weightings derive from (a) current attentional resources (as in [13]), and (b) the reinforcement history of the system, i.e., reward/punishment values assigned via personal and social exposure to cultural norms that dictate perspectives for understanding phenomena in context. In combination, these two factors determine that NL users exploit NL variability, not only for adjusting their understanding, but also for the purposes of attaching social/personal values to their actions, e.g., meanings signifying identity and group-memberships that go beyond denotational meaning (see e.g. [20]).

From the evolutionary point of view, variability is a constant property of a dynamically changing environment to which organisms have to actively respond controlling its influence moment by moment. To actively exert such control, an agent must store as part of its constitution predictable dependencies between its actions and the resulting sensory stimulation ('sensorimotor contingencies'). NL behaviour, as modelled by DS, is subsumed in this action resources organisation. Without imposing identity of strategies/outcomes at the level of individual agency, at the level of the social unit, co-construction of stimuli and meanings of the kind seen in split utterances (see (1)–(5) earlier) allows each interlocutor's processing to influence the other's actions by establishing feedback loops and thus lessening unpredictability leading towards temporary synergies of compatibility and coordination (see also [95]). For this fine-grained influence to take place, it is essential that what is sustaining the interactions is mechanisms flexibly shaping courses of actual/virtual actions (as in the DS DAG) rather than manipulation of stored fixed codes/intentions/goals/contents. Thus, on the DS perspective, the phenomenological phenomenon at some particular time in literate societies of a reified NL (code) can be seen as emerging from the high social values attached to stable (but, in fact, 'metastable' [55]) system states temporarily settling in short-term outcomes even though long-term the underlying basis is ephemeral ever-changing processes. Over the long term, by iterated interaction coordinations among groups of individuals, successful processing paths become progressively routinised and grammaticalised/lexicalised, i.e. easily activated as whole sequences of basic actions (*macros*). Cross-linguistic and diachronic analyses in DS show how the appearance of distinct NLs arises through the establishment of different such routinisations [7,9,59] invested with social value. This also provides the possibility of explanatory modelling of recent cognitive evidence that processing and interaction constraints affect directly the design of the grammar itself [28,77]. From this point of view, NLs can function adaptively, but also maladaptively in sedimenting prejudices and exclusions, because they comprise just mechanisms of storing and deploying reliable, systematic action-perception contingencies valid in particular human ecological niches, mainly the social environment. The perspective from which to develop the view of NLs as evolved systems is then broadly functionalist: narrowly defined, NL grammars store action-outcome contingencies for attracting and exploiting interaction with other humans (but also one's self in reasoning, planning, imagining) due to the reward values attached to interaction outcomes; in turn, such stored dispositions by prompting interaction and enabling it via the establishment of feedback loops and potential to aggregate as macros are shaped themselves to generalise efficiently due to iterative attempts at coordination with various partners and in various circumstances. Building on this basis, NL grammars can then become the underpinnings of systems of significance concerning moral/emotional considerations (e.g. altruism) and cultural group formations.

Though an adaptation-oriented view of NLs is advocated by some [11], the gradualist adaptation claim for NL grammars is disputed by others [30,64] invoking problems in reality caused by the code view of NLs. Despite marked differences between the various stand-points, two putative problems are assumed even in models taking the adaptionist perspective: (i) the problem of "signalling signalhood", i.e., identifying communicative intentions; and (ii) the assumption that successful communication requires establishing some fixed and shared signal-content correspondence ('compositionality') intended by the speaker to be recovered by the hearer. From the DS perspective, (i)–(ii) are artifacts of the reified code view of NLs. Regarding (i), "signalling signalhood" leads to definitional infinite regress and is only a consideration if a Gricean underpinning of communication is assumed under which the inadequacy of a code in this respect has to be supplemented by mindreading capacities. DS instead derives behaviour coordination via the NL mechanisms themselves, namely, the predictivity and adjustment of system resources. Any stimulus can be exploited not as an intentional 'signal' but as an affordance/conditional-action, depending on the current state of the agent. While retaining the assumptions of productivity and systematicity, DS rejects (ii), the standard compositionality requirement that imposes fixed NL form-content mappings. Instead, NLs are modelled as relatively reliable, but also fallible, processes, sets of domain-general basic procedures for licensing domain-specific action sequences that either assimilate input from the (social) environment (parsing) or induce behaviour to acquire that input (speech or other actions). Since it is not a reasonable assumption to impose duplication of needs and goals among individuals, variability of action-control mechanisms between conversational partners, or the same agent at different times, is expected. Moreover, fallibility is the main source of innovation, creativity, and increased efficiency [46] so this is not considered as an inherent problem that should have been eliminated by evolution. This is in line with evidence that in cross-generational acquisition, the evolution of underspecification/polysemy/ambiguity enhances rather than disadvantages language learning/change [60].

5.1 Language as an Adaptive Group-Creating Mechanism

In turning to evolutionary patterns, a distinctive feature of enactive approaches to the biology of living beings is that organisms are defined as autonomous systems that resist disorder and dissolution. Agents of various nested orders can be defined in this way, from cells, to individual brains, to dyads and groups, all repeatedly re-defining and configuring their own boundaries during interactions in ways that promote adaptive success and survival (see also [15,72]). NL coordination as modelled by DS is one such means of determining and maintaining boundaries ('group functional organisation', [92]), i.e., inducing temporary stabilities for joint action via social learning and cultural transmission.

Going beyond the level of the individual, the evolutionary concept of group selection has been resurrected by Sober and Wilson [82], who argue that groups must be treated as capable of constituting adaptive units in their own right, alongside individuals, for appropriate explanation of many evolutionary strands (the so-called *Multi-Level Selection Hypothesis*: [82,91]). Under this view, in most cases, group-level and individual-level pressures compete, with pressures for individual interests having to be outweighed by significant fitness enhancement at the level of competition among groups for group benefiting traits to be favoured. However, unlike many other macro-traits (e.g. moral behaviour), NL use is one of the few cases where both individual and group-level fitness seem to be affected given that NL abilities transcend the boundaries of individual vs social agency. Under the DS view, this is unsurprising since NLs are defined as means for interaction with interaction influencing each other's fitness being the very criterion that defines the concept of 'group' [82]. Instead of the standard notions of 'altruism' and 'cooperation' which omit the contribution of competition and individual vs social tensions as evolutionary forces, interaction in the DS sense can be seen as the basis of NL adaptivity (see also [70]). Moreover, given the abandonment of the code model, interaction and coordination between members of a group does not require that all members of the group share identical dispositions or intentions; as long as the propensities and goals of any interactants complement each other, they will be able to coordinate [66,69].

But we can also take a more fine-grained view, given that the concept of a 'group' is defined differentially relative to particular traits. Core NL features, namely, vagueness/ambiguity (i.e., open-endedness) and systematicity, have been shown to emerge solely from cultural transmission (iterative learning) without intentional design [60,78]. The challenge is to explain why this might be so. From this perspective, the relevance of individual- vs group-level distinctions in fitness enhancement arises. Firstly, a sufficiently loose concept of compositionality (*systematicity*) for affordance indicators allowing for variability in form and effects while nevertheless presuming on reliable predictable contributions is expected to arise at the individual level: individual memorization/learnability [60] needs a finite stock to systematic productive effect and, given that meaning is an emergent and relational feature of interactions, deterministic outcomes are not expected or needed. It is, however, cases where group adaptivity clearly outweighs individual adaptivity which provide the stronger evidence, solving what is otherwise puzzling, namely, that vagueness and ambiguity seem problematic at the individual level: open-endedness gives rise to the ever-present risk of misunderstanding between the interactive parties; and the related psychological correlate of non-determinism, uncertainty, phenomenologically seems hugely problematic for individuals. Nevertheless, the advantage of non-determinism at the group level is very striking. Open-endedness of action/perception outcomes via adaptable mechanisms is what enables group establishment. Individuals may interpret the world around them, including their interactions with other people, relative to their own needs and desires and still act collectively in coordination. But this can only be achieved if interactions between the parties do not demand identity of mechanisms or outcomes (see also [83]). Any such condition on achieving coordination would make NL communication appear impossible. In fact, it is the other way round, risk, uncertainty about even our own goals and resources, and vague offerings are the sine qua non of communication since meaning and innovation emerge through interaction, instead of being initially located in one mind and having to be transferred to another. Successful employment of only apparently shared terms in the service of variable purposes/meanings is then explainable if the assumption is not made that prior "common ground" and subsequent duplication of contents is a necessary presupposition for joint action. Here history provides numerous examples of how the unifying force of NL terms across disjoint communities can achieve striking social success through enabling otherwise conflicting sub-groups to cooperate under a single label; examples include *Solidarity* of Poland, *Coordinadora* in Bolivia, and the Zapatista rising in Mexico [85].

6 Conclusion

The DS perspective aims to directly model the group-forming properties of NL interactions. First, NL communication is not viewed as convergence/common ground but as the employment of procedures enabling creative joint activities without overarching common goals. Secondly, we have barely scratched the surface of a great number of issues here only perhaps to argue sufficiently that it is notable that a common non-individualistic pattern can be discerned across NL learning, individual and institutional NL change, and evolution. At all stages, modelling relies in situated iteration: the entrenching effect of assigning higher probability weightings to iterated processing paths (given DS assumptions) leading to routinisation; the setting up of shortcuts in response to cognitive pressures for economy; all being buttressed by the group functional organisation which the interactivity induces. NL learning, change and even NL emergence can all then be seen in gradualistic terms, hence the higher-order organisation that incorporates the NL system itself can be argued to constitute an adaptive interactive system in continuity with the definition of living organisms as modelled in enactive approaches.

References

1. Altmann, G., Kamide, Y.: Incremental interpretation at verbs: restricting the domain of subsequent reference. Cognition **73**(3), 247–264 (1999)
2. Anderson, M.L.: After Phrenology. Cambridge University Press, Cambridge (2014)
3. Anderson, M.L.: Of Bayes and Bullets: An Embodied, Situated, Targeting-Based Account of Predictive Processing. Johannes Gutenberg-Universität, Mainz (2017)
4. Arnon, I., Casillas, M., Kurumada, C., Estigarribia, B.: Language in interaction: Studies in honor of Eve V Clark, vol. 12. John Benjamins, Amsterdam (2014)
5. Berwick, R.C., Chomsky, N.: Why Only Us: Language and Evolution. MIT, London (2015)
6. Bickhard, M.H., Richie, D.M.: On the Nature of Representation: A Case Study of James Gibson's Theory of Perception. Praeger, New York (1983)
7. Bouzouita, M.: The Diachronic Development of Spanish Clitic Placement. Ph.D. thesis, King's College London (2008)
8. Cann, R., Kempson, R., Marten, L.: The Dynamics of Language. Elsevier, Oxford (2005)
9. Chatzikyriakidis, S.: Clitics in Grecia Salentina Greek: a dynamic account. Lingua **119**(12), 1939–1968 (2009)
10. Chomsky, N.: Knowledge of Language: Its Nature, Origin, and Use. Praeger, Santa Barbara (1986)
11. Christiansen, M.H., Chater, N.: Language as shaped by the brain. Behav. Brain Sci. **31**(5), 489–509 (2008)
12. Cisek, P.: Cortical mechanisms of action selection: the affordance competition hypothesis. Phil. Trans. R. Soc. B: Bio. Sci. **362**(1485), 1585–1599 (2007)
13. Clark, A.: Surfing Uncertainty: Prediction, Action, and the Embodied Mind. Oxford University Press, Oxford (2016)
14. Clark, A.: Busting out: predictive brains, embodied minds, and the puzzle of the evidentiary veil. Noûs **51**(4), 727–753 (2017)
15. Clark, A.: How to Knit Your Own Markov blanket. Johannes Gutenberg-Universität Mainz, Germany (2017)
16. Clark, E.V., Casillas, M.: First language acquisition. In: Allan, K. (ed.) The Routledge Handbook of Linguistics, pp. 311–329. Routledge, London (2016)
17. Clark, H.H.: Communal lexicons. In: Malmkjær, K., Williams, J. (eds.) Context in Language Learning and Language Understanding, vol. 4, pp. 63–87. Cambridge University Press, Cambridge (1998)
18. Cooper, R.: Type theory and semantics in flux. In: Kempson, R., Asher, N., Fernando, T. (eds.) Handbook of the Philosophy of Science. Philosophy of Linguistics, vol. 14, pp. 271–323. North Holland, Oxford (2012)
19. Dingemanse, M., Roberts, S.G., Baranova, J., Blythe, J., Drew, P., et al.: Universal principles in the repair of communication problems. PloS One **10**(9), e0136100 (2015)
20. Eckert, P.: Three waves of variation study: the emergence of meaning in the study of sociolinguistic variation. Annu. Rev. Anthropol. **41**, 87–100 (2012)
21. Eshghi, A., Howes, C., Gregoromichelaki, E., Hough, J., Purver, M.: Feedback in conversation as incremental semantic update. In: Proceedings of the 11th IWCS, pp. 261–271. ACL (2015)
22. Eshghi, A., Purver, M., Hough, J.: DyLan: Parser for Dynamic Syntax. Technical Report of Queen Mary University, London (2011)

23. Eshghi, A.: DS-TTR: an incremental, semantic, contextual parser for dialogue. In: Proceedings of the 19th SemDial (2015)
24. Eshghi, A., Lemon, O.: How domain-general can we be? Learning incremental dialogue systems without dialogue acts. In: Proceedings of the 18th SemDial (2014)
25. Eshghi, A., Purver, M., Hough, J.: Probabilistic induction for an incremental semantic grammar. In: Proceedings of the 10th IWCS, pp. 107–118. Potsdam (2013)
26. Eshghi, A., Purver, M., Hough, J., Sato, Y.: Inducing lexical entries in an incremental semantic grammar. In: Proceedings of CSLP2012 (2012)
27. Eshghi, A., Shalyminov, I., Lemon, O.: Interactional dynamics and the emergence of language games. In: Proceedings of FADLI Workshop, ESSLLI (2017)
28. Fedzechkina, M., Chu, B., Florian Jaeger, T.: Human information processing shapes language change. Psychol. Sci. **29**(1), 72–82 (2017). https://doi.org/10.1177/0956797617728726
29. Ferreira, F., Chantavarin, S.: Integration and prediction in languageprocessing: a synthesis of old and new. Curr. Dir. Psychol. Sci. **27**(6), 443–448 (2018)
30. Fitch, W.T., Hauser, M.D., Chomsky, N.: The evolution of the language faculty: clarifications and implications. Cognition **97**(2), 179–210 (2005)
31. Friston, K., Frith, C.: A duet for one. Conscious. Cogn. **36**, 390–405 (2015)
32. Gallagher, S.: Enactivist Interventions: Rethinking the Mind. Oxford University Press, Oxford (2017)
33. Gargett, A., Gregoromichelaki, E., Howes, C., Sato, Y.: Dialogue-grammar correspondence in dynamic syntax. In: Proceedings of the 12th SemDial (2008)
34. Gargett, A., Gregoromichelaki, E., Kempson, R., Purver, M., Sato, Y.: Grammar resources for modelling dialogue dynamically. Cogn. Neurodyn. **3**(4), 347–363 (2009)
35. Gibson, J.J.: The Ecological Approach to Visual Perception. Psychology Press, Oxford (1979)
36. Ginzburg, J.: The Interactive Stance. OUP, Oxford (2012)
37. Gregoromichelaki, E.: Conditionals: a dynamic syntax account. Ph.D. thesis, King's College London (2006)
38. Gregoromichelaki, E.: Grammar as action in language and music. In: Orwin, M., Kempson, R., Howes, C. (eds.) Language, Music and Interaction, pp. 93–134. College Publications, London (2013)
39. Gregoromichelaki, E.: Quotation in dialogue. In: Saka, P., Johnson, M. (eds.) The Semantics and Pragmatics of Quotation. Perspectives in Pragmatics, Philosophy & Psychology, vol. 15. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-68747-6_8
40. Gregoromichelaki, E., Kempson, R.: Joint utterances and the (split-) turn taking puzzle. In: Mey, J.L., Capone, A. (eds.) Interdisciplinary Studies in Pragmatics, Culture and Society. Perspectives in Pragmatics, Philosophy & Psychology, vol. 4. Springer, Heidelberg (2016). https://doi.org/10.1007/978-3-319-12616-6_28
41. Gregoromichelaki, E., Kempson, R., Howes, C., Eshghi, A.: On making syntax dynamic. In: Wachsmuth, I., de Ruiter, J.P., Jaecks, P., Kopp, S. (eds.) Alignment in Communication, pp. 57–85. John Benjamins, Amsterdam (2013)
42. Gregoromichelaki, E., Kempson, R., Purver, M., Mills, G.J., Cann, R., Meyer-Viol, W., Healey, P.G.T.: Incrementality and intention-recognition in utterance processing. Dialogue Discourse **2**(1), 199–233 (2011)
43. Gregoromichelaki, E., et al.: Completability vs (in)completeness. Acta Linguistics Hafniensia (forthcoming)

44. Gregoromichelaki, E., Sato, Y., Kempson, R., Gargett, A., Howes, C.: Dialogue modelling and the remit of core grammar. In: Proceedings of IWCS (2009)
45. Healey, P.G.T., Purver, M., Howes, C.: Divergence in dialogue. PLoS ONE **9**(6), 98598 (2014)
46. Healey, P.G.T., de Ruiter, J.P., Mills, G.J.: Editors' introduction: miscommunication. Top. Cogn. Sci. **10**(2), 264–278 (2018)
47. Hendriks, P.: Asymmetries Between Language Production and Comprehension. Studies in Theoretical Psycholinguistics, vol. 42, 1st edn. Springer, Dordrecht (2014). https://doi.org/10.1007/978-94-007-6901-4
48. Hilbrink, E.E., Gattis, M., Levinson, S.C.: Early developmental changes in the timing of turn-taking: a longitudinal study of mother-infant interaction. Front. Psychol. **6**, 1492–1492 (2015)
49. Hough, J.: Modelling incremental self-repair processing in dialogue. Ph.D. thesis, Queen Mary University of London (2014)
50. Hough, J., Purver, M.: Probabilistic type theory for incremental dialogue processing. In: 2014 Proceedings of the EACL, pp. 80–88. ACL, April 2014
51. Howes, C., Eshghi, A.: Feedback relevance spaces. In: 2017 Proceedings of the 12th International Conference on Computational Semantics IWCS (2017)
52. Howes, C., Healey, P.G.T., Purver, M., Eshghi, A.: Finishing each other's ... responding to incomplete contributions in dialogue. In: Proceedings of the 34th Annual Meeting of the Cognitive Science Society, pp. 479–484 (2012)
53. Hurley, S.: The shared circuits model. Behav. Brain Sci. **31**(1), 1–22 (2008)
54. Kalatzis, D., Eshghi, A., Lemon, O.: Bootstrapping incremental dialogue systems. In: Proceedings of Conference on Neural Information Processing Systems (2016)
55. Kelso, J.S.: Multistability and metastability: understanding dynamic coordination in the brain. Phil. Trans. R. Soc. B **367**(1591), 906–918 (2012)
56. Kempson, R., Cann, R., Gregoromichelaki, E., Chatzikyriakidis, S.: Language as mechanisms for interaction. Theor. Linguist. **42**(3–4), 203–276 (2016)
57. Kempson, R., Gregoromichelaki, E.: Action sequences instead of representational-levels. Behav. Brain Sci. **40**, 296 (2017)
58. Kempson, R., Meyer-Viol, W., Gabbay, D.: Dynamic Syntax: The Flow of Language Understanding. Blackwell, Oxford (2001)
59. Kiaer, J.: Processing and interfaces in syntactic theory: the case of Korean. Ph.D. thesis, King's College London (2007)
60. Kirby, S., Smith, K., Cornish, H.: Language, learning and cultural evolution. In: Cooper, R., Kempson, R. (eds.) Language in Flux, pp. 81–108. College Publications, London (2008)
61. Kuperberg, G.R., Jaeger, T.F.: What do we mean by prediction in language comprehension? Lang. Cogn. Neurosci. **31**(1), 32–59 (2016)
62. Lerner, G.H.: Collaborative turn sequences. In: Conversation Analysis: Studies from the First Generation, pp. 225–256. John Benjamins, Amsterdam (2004)
63. Lerner, G.H.: On the place of linguistic resources in the organization of talk-in-interaction: grammar as action in prompting a speaker to elaborate. Res. Lang. Soc. Interact. **37**(2), 151–184 (2004)
64. Lightfoot, D.: The Development of Language. Blackwell, Oxford (1999)
65. Lowder, M.W., Ferreira, F.: Prediction in the processing of repair disfluencies. J. Exp. Psychol.: Learn. Mem. Cogn. **42**(9), 1400–1416 (2016)
66. Millikan, R.: The Varieties of Meaning: The Jean-Nicod Lectures. MIT Press, Cambridge (2004)
67. Millikan, R.G.: On Clear and Confused Ideas. Cambridge University Press, Cambridge (2000)

68. Mills, G., Gregoromichelaki, E.: Establishing coherence in dialogue: sequentiality, intentions and negotiation. In: Proceedings of the 14th SemDial (2010)
69. Mills, G.J.: Dialogue in joint activity. New Ideas Psychol. **32**, 158–173 (2014)
70. Nettle, D., Dunbar, R.I.M.: Social markers and the evolution of reciprocal exchange. Curr. Anthropol. **38**(1), 93–99 (1997)
71. Odling-Smee, F.J., Laland, K.N., Feldman, M.W.: Niche Construction: The Neglected Process in Evolution, vol. 37. Princeton University Press, Princeton (2003)
72. Paolo, E.A.D., Cuffari, E.C., Jaegher, H.D.: Linguistic Bodies. MIT Press, Cambridge (2018)
73. Piaget, J.: The Construction of Reality in the Child. Basic Books, New York (1954)
74. Pickering, M.J., Ferreira, V.S.: Structural priming: a critical review. Psychol. Bull. **134**(3), 427 (2008)
75. Pickering, M.J., Garrod, S.: An integrated theory of language production and comprehension. Behav. Brain Sci. **36**, 329–347 (2013)
76. Purver, M., Gregoromichelaki, E., Meyer-Viol, W., Cann, R.: Splitting the 'I's and crossing the 'You's. In: Proceedings of the 14th SemDial, pp. 43–50 (2010)
77. Roberts, S.G., Levinson, S.C.: Conversation, cognition and cultural evolution. Interact. Stud. **18**(3), 402–429 (2017)
78. Santana, C.: Ambiguity in cooperative signaling. Philos. Sci. **81**(3), 398–422 (2014)
79. Sato, Y.: Local ambiguity, search strategies and parsing in dynamic syntax. In: Gregoromichelaki, E., Kempson, R., Howes, C. (eds.) The Dynamics of Lexical Interfaces. CSLI Publications, Stanford (2011)
80. Saxton, M.: Child Language. Sage, Los Angeles (2017)
81. Shalyminov, I., Eshghi, A., Lemon, O.: Challenging neural dialogue models with natural data. In: Proceedings of the 21st SemDial. pp. 98–106 (2017)
82. Sober, E., Wilson, D.S.: Unto Others, vol. 218. Harvard University Press, Cambridge (1998)
83. Sowa, J.F.: Language games. In: Pietarinen, A.V. (ed.) Game Theory and Linguistic Meaning, pp. 17–38. Elsevier, Amsterdam (2007)
84. Sperber, D., Wilson, D.: Relevance, 2nd edn. Blackwell, Oxford (1995)
85. Tarrow, S.: Power in Movement. Cambridge University Press, Cambridge (2011)
86. Thompson, E.: Mind in Life. Harvard University Press, Cambridge (2007)
87. Tomasello, M.: Origins of Human Communication. MIT Press, Cambridge (2008)
88. Trevarthen, C.: The child's need to learn a culture. Child. Soc. **9**(1), 5–19 (1995)
89. Trueswell, J.C., Tanenhaus, M.K.: Approaches to Studying World-Situated Language Use. MIT Press, Cambridge (2005)
90. Vygotski, L.S.: Thought and Language. MIT Press, Cambridge (2012)
91. Wilson, D.S.: Darwin's Cathedral. University of Chicago Press, Chicago (2002)
92. Wilson, D.S.: Does Altruism Exist?: Culture, Genes, and the Welfare of Others. Yale University Press, New Haven (2015)
93. Yu, Y., Eshghi, A., Lemon, O.: Incremental generation of visually grounded language in situated dialogue. In: Proceedings of INLG, Los Angeles (2016)
94. Yu, Y., Eshghi, A., Lemon, O.: Learning how to learn: an adaptive dialogue agent for incrementally learning visually grounded word meanings. In: RoboNLP (2017)
95. Zawidzki, T.W.: Mindshaping. MIT Press, Cambridge (2013)
96. Zeevat, H.: Language Production and Interpretation. Brill, Frankfurt (2014)

Author Queries

Chapter 11

Query Refs.	Details Required	Author's response
AQ1	Per Springer style, both city and country names must be present in the affiliations. Accordingly, we have inserted the city and country names in affiliations. Please check and confirm if the inserted city and country names are correct. If not, please provide us with the correct city and country names.	
AQ2	This is to inform you that corresponding author has been identified as per the information available in the Copyright form.	