Cognitive science, language as a tool for interaction, and a new look at language evolution

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

Stergios Chatzikyriakidis & Christine Howes
Philosophy, Linguistics and Theory of Science
University of Gothenburg
{stergios.chatzikyriakidis, christine.howes}@gu.se

Abstract

We explore prerequisites necessary for embedding Dynamic Syntax within an account of language evolution. We show how the dynamics of processing as modelled in Dynamic Syntax display remarkable parallelism with Clark's (2016) Predictive Processing Model and that the interactive stance of a combined DS/PPM model of language/cognition reflects the Multi-Level Selection Hypothesis – with groups as units for evolutionary purposes, not just individuals. With these assumptions, language emerges without necessary invocation of rich innate encapsulated structures or mind-reading capacities, paralleling first language acquisition.

1 Introduction

This paper sets out a new direction for language evolution research that brings together three disciplines in a novel cross-disciplinary perspective.

2 Dynamic Syntax

The starting point is from within Linguistics, and the Dynamic Syntax stance (DS: Cann et al., 2005) in which the grammar framework itself sets out the dynamics of how the information growth process is achieved in both parsing and production (Kempson et al., 2016). This tight, system-internal integration is strikingly made evident by informal dialogue exchanges in which speakers and hearers jointly induce structure, fluently and effortlessly switching roles at arbitrary points in an exchange (1)-(9). Interlocutors can each contribute a fragment (1) with the overall content and inferential effect only emerging across parties in a composite group activity. Although these fragments are structurally collaborative, this does not necessitate the recovery of some previously or even subsequently agreed intended propositional content or speech-act (1,8): notably, even very young children are able to join in with appropriate increments in the joint activity of co-creating dialogue (5,6), well before evidence of productive mind-reading capability (Breheny, 2006).

This universal phenomenon, i.e. *split utterances*, is highly problematic for all major grammatical frameworks, with their exclusive emphasis on licensing sentence-string/interpretation pairings as output of the grammar system. As these are taken to model the ideal speaker/hearer's capacity in language, all split utterance data are beyond their remit. They are largely ignored and/or explained as performance disfluencies. This judgement, however, flies in the face of the seamless fluency of the phenomenon in informal conversation. DS apart (Gregoromichelaki et al., 2011, a.o.), the only studies addressing this challenge (Poesio and Rieser, 2010; Pickering and Garrod, 2013) are at best incomplete in modelling only the subset of deliberately 'helpful' completions (2,3); and such accounts involve complex externally imposed operations: e.g. high-level inference (Gibson et al., 2013), abduction (Friston and Frith, 2015) or the creation of efference-copies as in standard models of action control (e.g. Pickering and Garrod, 2013; though see Clark, 2016). Since very young children freely join in on such utterance exchanges, the full complexity of any such mechanisms has to be assumed to be in place prior to the acquisition process if these are to be captured, a view notably embraced in its strongest form by Tomasello and colleagues (2005; 2008) in the form of innate specification of Gricean inference capacities.

In DS, no such commitment is necessary, as the split utterance effects follow directly from the system itself. Production and parsing both involve the top-down anticipation-driven construction of al-

(1) **A:** We're going to ... **B:** Burbage to see Ann, Auntie Ann **C:** with the dogs? **B:** if you look after them.	(2) **Homeowner:** I shall need the ... **Gardener:** mattock. For breaking up clods of earth. [BNC]	(3) **A:** Have all the students handed in **B:** their term papers? **A:** or even any assignments
(4) **Sue:** I'm afraid I burned the kitchen ceiling. **Michael:** Did you burn **Sue:** myself? No, fortunately not.	(5) **Carer:** Old McDonald had a farm... E-I-E-I-O. And on that farm he had a... **Child:** cow.	(6) **Teacher:** And your name is ... **Child:** Mary
(7) **A:** And they ignored the conspirators who were **B:** Geoff Hoon and Patricia Hewitt [BBC Radio 4 06/09/10]	(8) (A and B arguing:) **A:** It's clear from what you've just said that **B:** that I am completely vindicated	(9) (T(herapist) and C(lient)) **T:** Your sponsor before... **C:** was a woman. **T:** Yeah. **C:** But I only called her every three months. **T:** And so your sobriety now, in AA [is] **C:** [is] at a year [Ferrara 1992]

Figure 1: Examples

ternative interpretations directly establishing step by step coordination, with competing emergent interpretations involving probabilistic weightings with consistency checking filtering out errors (Eshghi et al., 2013; Hough and Purver, 2014; Kempson et al., 2015), and positive and negative feedback constraining the searchspace (Eshghi et al., 2015). Online processing is thus modelled as system-internal structural growth (Chatzikyriakidis and Kempson, 2011; Kempson et al., 2016) and not via the grammar plus externally defined parsing/production modules. This captures standard sentence or subsentential-level morphosyntactic phenomena, all the way up to discourse effects such as ellipsis and dialogue data (1)-(9). This is an advantage over other frameworks, in which some (sometimes non-conservative) extensions need to be made.

On the DS view, to the contrary, the interaction emerges from the fact that all parties are using the same structure building strategies. The coordinative effect is a direct consequence of incorporating the dynamics of online processing within the grammar formalism. There is no invocation or presumption of grammar-external inference to achieve the interactivity intrinsic to dialogic exchanges. However, although this ability is freely made use of once it has become available to the language user, it is not a sine qua non for language development. Indeed, even on the assumption that a mind-reading capacity plays a large part in adult cooperativity, it is notable that the assumptions of shared IDENTITY are never guaranteeable. Despite this, the effectiveness of interaction is almost never jeopardised, and even in cases where corrections/clarifications are warranted, successful interactive exchange is buttressed through such overt clarifications.

Recent computational work confirms both the viability of DS as a grammar formalism, its fit to include multimodal data to parse or generate complex dialogue data (e.g. corrections, elliptical fragments, split utterances), and the provision it makes available to enable a large amount of dialogue data to be acquired from very small amounts of unannotated data (actually, just one sentence), using a combination of DS and Reinforcement Learning (Yu et al., 2016; Kalatzis et al., 2016). Earlier work (Eshghi et al., 2013) has shown that an incremental semantic grammar can be acquired using limited data. Both the training and test data are taken from utterances in the CHILDES corpus paired with their logical forms expressed using Type Theory with Records (TTR: Cooper, 2005). The system input is comprised of: (i) a fixed set of computational DS actions (language general structure building mechanisms); (ii) a training set of the form $<S_i, RT_i>$, where $<S_i>$ are sentences of the language and $<RT_i>$ their targeted semantic representations. The output induces lexical actions for the individual words, probabilistically decomposing the possible sequences of actions that lead to the complete target semantic representations.[1] The results suggest that grammar induction of a probabilistic grammar in an incremental, semantic model like DS can be done effectively without prior assumptions of syntactic structure. On a more general level, this might point to the possibility of a set of language independent (and potentially domain general) computational actions that given appropriate data can induce domain specific systems (e.g. lexical actions for individual words).

3 The Predictive Processing Model

This perspective fits directly into a larger cognitive science perspective along two dimensions, that of

[1]The interested reader is directed to Eshghi et al. (2013) for more details.

an integrated nonmodular cognitive system, and that of how language might have evolved as a departure within such a cognitive system. First, we find notable parallels between the Dynamic Syntax perspective and the Clark (2013; 2016) view of cognition modelled as a generative Predictive Process (PPM). PPM equally argues that action and perception act in tandem and invokes neither higher order intentions nor an efference copy construction to yield interactional effects. In PPM, as in DS, context is everything: brains are predictive engines (not passive modular input systems) using their own immediate and encyclopaedic context at every step to guess the structure/shape of the incoming sensory array, which forms 80% of the burden of processing. DS can thus be viewed as a specialisation of the dynamics of the PPM framework, viewing language as a set of action transitions from one context state to another. Context itself is also characterised in process-terms of growth rather than a static store. The addition of DS extends the PPM by adding the dimension of manifest interaction for which language is the central tool. Within this extended model, there are two variants; and on either variant, this nesting of the DS model of language within PPM yields a yet larger perspective – that of language evolution.

Current work on language evolution is split between two extremes. At one end are those who see language as innate and encapsulated. Language is just one amongst a (large) number of modules, each with their specialised niche, requiring some form of glue-language to relate one vocabulary with another; and no adaptationist account is possible (Fitch et al., 2005). At the other extreme are those who see acquisition of language as emerging out of the dynamics of communication, with the panoply of Gricean-style axioms and mind-reading capacities taken to be a necessary prerequisite for acquisition of language, hence innate (Tomasello, 2008; Jaeger, 2007; Christiansen and Chater, 2016). Under all these views, the emergence of language is parasitic on the not unproblematic assumption that language itself is a specialised module, not reducible to more general cognitive architecture. Either type of account requires a significant shift away from a general inferentialist system by some form of switch mechanism to an encapsulated language faculty not expressible in the same terms as the general cognitive system. There are variants between these extremes, amongst which Kirby et al. (2008) argue that compositionality in language can be learned without predispositions, offering a counter-argument to the innateness view of language and its anti-adaptationist stance.

4 The Multilevel Selection Hypothesis

In addition, a combined DS/PPM perspective suggests a view which reflects recent work in evolutionary biology urging a re-evaluation of groups as a unit for evolutionary purposes: the Multilevel Selection Hypothesis (MLSH: Sober and Wilson 1998; Wilson, 2002). On the MLSH view, evolution is seen as driven by two separate dimensions, individual- and group- level adaptivity. The potential of a group to form an adaptive unit turns on the successful balancing of these two conflicting dimensions requiring intra-group pressure to moderate rampant individualism. This dual-level perspective has not so far been taken up within the language evolution narrative, which remains based on individualistic competing selfish considerations. We now explore this in two steps.

Taking first an individual-centred basis DS offers a view of language broadly following (Kirby et al., 2008) in not having to stipulate either rich innate attributes of structure, or externally imposed innate higher-order inference capabilities as pre-requisite to language development, while opening up the potential for an MLSH form of explanation. With DS assumptions, the interactivity displayed by split utterances is seen as emerging from a background of rich interaction between co-participants without any necessity of shared agreed content, as vividly displayed in first language acquisition (Hilbrink et al., 2015). The in tandem co-construction by speaker and hearer of some sound-interpretation pairing is grounded in the already robustly established pattern of situated interactional behaviour between carer and child. The infant's non-language-based verbalising behaviour is interpreted by the carer as contributing to some verbal frame which she herself may have as the basis for engaging with the child in order to create the bonding achievable – even without any signalled content being conveyed (e.g. the peekaboo games which pre-linguistic children so enjoy; Clark and Casillas, 2016). Fragments such as one word utterances initiating the child's emerging language capability are also interpreted against the rich contextualisation of the carer, either in interpreting the child's minimal utterance, or in providing a frame relative to which the utterance provides an entirely successful completion (as in (5,6)), building on the pleasure in interaction which the infant and carer already share.

It is then a small second step to see this established interactivity as the basis for a new group-oriented perspective on language evolution. Successful utterance exchanges, even one word utterances, can be seen as achieving the same context-dependent interactional effect displayed by other primates but with the addition of manifest signalling of that interactional effect – from which the step of ascribing content to a signal could have developed (Kirby et al., 2008; Scott-Phillips et al., 2009). The inexorable interactive duplication by all parties in jointly building up the substructure to meaningfully support such utterances yields cumulative interactive effects, multiplied recursively with each additional language token. And with such interactions, repeated reiterations combine with internal cognitive pressure for simplification and cognitive economy and inexorably lead to routinisation effects, with macro sequences of actions becoming stored for ease of recoverability. This leads to recursive buttressing of the group ethos, without ever needing the identity of word tokens or their interpretation to be manifestly confirmed. Hence the uncontentiously effective group-forming trait of language which creates sharp barriers against those who cannot control the stored routinised string-interpretation pairings necessary to achieve the interactiveness that the language makes manifest. Moreover, though the role of mind-reading and explicit seeking of common goals in later stages of language and cognitive development remains an undoubted buttressing force for group consolidation, it no longer plays a role in triggering language emergence: rather, the merely approximate cross-speaker correspondence of string-interpretations set by each participant contributes to gradual language change in the face of cognitive and social pressures.

This contrasts with the Tomasello (2008) account in particular, which claims that the full apparatus of Gricean reasoning has to be innate: "communicative intentions of the cooperative (Gricean) kind [are] clearly a prerequisite for understanding symbols"; and "the idea of language without shared intentionality, even in one-unit expressions, is simply incoherent."(Tomasello et al., 2005, 724). It is notable that the considerable empirical data supposedly confirming this innate cooperativity and desire to be helpful to others, claimed to need dual representation of both speaker and hearer perspective for each individual participant, can all be explained relative to the weaker stance that it is the potential for interaction which is innate, and not a necessity of "shared contents" or "shared goals", with their problematic concept of identity of content. Tomasello et al. (2005) note the potential functionality of shared intentionality at the level of group selection, but do not develop it. On the DS view, the group dynamic IS the story, irreducibly so, in virtue of the characterisation of language as manifest mechanisms for securing on-line interactive exchange.

Finally, we can now see how the evolutionary advantage of language lies in its adaptivity both at the individual and group level. In contradistinction to both biological and cultural evolution (Sober and Wilson, 1998; Wilson, 2002), in which selfish behaviour is seen as having to be kept within bounds if optimal adaptivity is to be ensured, capacity for language is advantageous both for individual- and group- level adaptivity. It is adaptive for the individual because it enhances potential for interactive and cooperative exchange with others, with individual benefits in such co-operation not achievable without language. It is adaptive for the group because it buttresses group potential for survival in accentuating and distinguishing other competing groups. The claimed relative adaptivity of individual languages in their progressive shift to meeting the brain desiderata of providing input able to be processed fast against ever-evolving contexts (Christiansen and Chater, 2016), can now be replaced with the more appropriate view of languages as differing in the various culturally evolved sets of actions they license, all of them being subject to cognitive constraints associated with the pressures for rapid real-time processing.

References

Richard Breheny. 2006. Communication and folk psychology. *Mind & Language*, 21(1):74–107.

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

Stergios Chatzikyriakidis and Ruth Kempson. 2011. Standard modern and pontic greek person restrictions: A feature-free dynamic account. *Journal of Greek Lingusitics*, pages 127–166.

Morton H. Christiansen and Nick Chater. 2016. *Creating Language: Integrating Evolution, Acquisition, and Processing*. MIT Press.

Eve V. Clark and M. Casillas. 2016. First language acquisition. In Keith Allan, editor, *The Routledge Handbook of Linguistics*, pages 311–329. Routledge.

Andy Clark. 2013. Are we predictive engines? perils, prospects, and the puzzle of the porous perceiver. *Behavioral and Brain Sciences*, 36(03):233–253.

Andy Clark. 2016. *Surfing uncertainty: Prediction, Action, and the Embodied Mind*. Oxford University Press.

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

Arash Eshghi, Julian Hough, and Matthew Purver. 2013. Incremental grammar induction from child-directed dialogue utterances. In *Proceedings of the 4th Annual Workshop on Cognitive Modeling and Computational Linguistics*, pages 94–103. ACL.

Arash Eshghi, Christine Howes, Eleni Gregoromichelaki, Julian Hough, and Matthew Purver. 2015. Feedback in conversation as incremental semantic update. In *Proceedings of the 11th International Conference on Computational Linguistics IWCS*.

Tecumseh Fitch, Mark Hauser, and Noam Chomsky. 2005. The evolution of the language faculty: clarifications and implication. *Cognition*, 97:197–201.

Karl Friston and Christopher Frith. 2015. A duet for one. *Consciousness and cognition*, 36:390–405.

Edward Gibson, Steven T Piantadosi, Kimberly Brink, Leon Bergen, Eunice Lim, and Rebecca Saxe. 2013. A noisy-channel account of crosslinguistic word-order variation. *Psychological Science*.

Eleni Gregoromichelaki, Ruth Kempson, Matthew Purver, Greg J. Mills, Ronnie Cann, Wilfried Meyer-Viol, and Patrick G. T. Healey. 2011. Incrementality and intention-recognition in utterance processing. *Dialogue and Discourse*, 2(1):199–233.

Elma E. Hilbrink, Merideth Gattis, and Stephen C. Levinson. 2015. Early developmental changes in the timing of turn-taking: a longitudinal study of motherinfant interaction. *Frontiers in Psychology*, 6:1492.

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

Gerhard Jaeger. 2007. Evolutionary game theory and typology: A case study. *Language*, (3(1)):74–109.

Dimitrios Kalatzis, Arash Eshghi, and Oliver Lemon. 2016. Bootstrapping incremental dialogue systems: using linguistic knowledge to learn from minimal data. In *Proceedings of the NIPS 2016 workshop on Learning Methods for Dialogue*, Barcelona.

Ruth Kempson, Eleni Gregoromichelaki, Arash Eshghi, and Julian Hough. 2015. Ellipsis in dynamic syntax. In Jeroen van Craenenbroeck and Tanja Temmerman, editors, *The Oxford Handbook of Ellipsis*. Oxford University Press.

Ruth Kempson, Ronnie Cann, Eleni Gregoromichelaki, and Stergios Chatzikyriakidis. 2016. Language as Mechanisms for Interaction. *Theoretical Linguistics*, 42(3-4):203–276.

Simon Kirby, Kenny Smith, and Hannah Cornish. 2008. Language, learning and cultural evolution. In Robin Cooper & Ruth Kempson, editor, *Language in Flux: dialogue coordination, language variation, change and evolution*, pages 81–108. College Pubs.

Martin J. Pickering and Simon Garrod. 2013. An integrated theory of language production and comprehension. *Behavioral and Brain Sciences*, 36:329–347, 8.

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

Thomas C. Scott-Phillips, Simon Kirby, and Graham R. S. Ritchie. 2009. Signalling signalhood and the emergence of communication. *Cognition*, 113(2):226–233.

Elliott Sober and David Sloan Wilson. 1998. *Unto others: The evolution and psychology of unselfish behavior*. Number 218. Harvard University Press.

Michael Tomasello, Malinda Carpenter, Josep Call, Tanya Behne, and Henrike Moll. 2005. Understanding and sharing intentions: The origins of cultural cognition. *Behavioral and Brain Sciences*, 28:675–691.

Michael Tomasello. 2008. *Origins of Human Communication*. MIT press.

David Sloan Wilson. 2002. *Darwin's cathedral: Evolution, religion, and the nature of society*. University of Chicago Press.

Yanchao Yu, Arash Eshghi, and Oliver Lemon. 2016. Incremental generation of visually grounded language in situated dialogue. In *Proceedings of INLG 2016*, Los Angeles.