DE GRUYTER MOUTON
DOI 10.1515/langcog-2013-0012 — Language and Cognition 2013; 5(2–3): 177–188

Peter Ford Dominey

The tip of the language iceberg

Abstract: Arbib's How the brain got language is a major achievement in defining a trajectory for the evolution of complex imitation and the language-ready brain leading to human language. In addition to these capabilities, I will suggest that it is useful to consider two additional components of human brain function that are intricately related to the emergence of language. These are, first, the profound human motivation to represent and share the psychological states of others, and second, the related complex semantic system that represents the contents of what is communicated in language. In this sense, these two components represent part of what is under the iceberg, where language is the emerging tip.

Keywords: social interaction, semantics

Peter Ford Dominey: Robot Cognition Laboratory, INSERM U846 Stem Cell and Brain Research Institute, 69675 Bron, France. E-mail: Peter.dominey@inserm.fr

1 Overview

How the Brain Got Language is a timely work that provides a thorough guide through important aspects of the evolutionary physiological preconditions for the development of language. This is a commendable effort, and the result is an important characterization of how the language ready brain got that way. Of course one cannot do everything in one book. This comment addresses two areas that can be improved.

The first aspect concerns the uniquely human drive to share psychological states with others and unique underlying forms of cognitive representation for doing so, as illustrated in the research of Tomasello and colleagues (Tomasello et al. 2005). This includes both the behavior itself, i.e. the motivation to share mental states, as well as the underlying neurophysiology. One goal of the comment is to indicate how this social dimension can be integrated in a manner that is compatible and complimentary with Arbib's approach. This raises a question as to how the language-ready brain became equipped for intentional communication. This in turn leads to a question concerning the uniquely human motivation to share mental states. Interestingly, Finlay (Syal and Finlay 2011) has suggested that linking strong motivational inputs to circuits related to perceptual and motor aspects of communicative signaling (including vocalization, gaze and facial expression) could have contributed to the development of socially motivated communication.

The second area that this comment will address is related to the construction grammar framework that is adopted, and some resulting issues related to compositionality in semantics and syntax.

2 Motivation to communicate

Arbib develops the mirror neuron hypothesis by which one can interpret other's actions, and the associated extension of this system into the communicative domain, providing a stepwise characterization of the trajectory necessary for the evolution of a language-ready brain. However, there is something missing. The missing element reveals itself even in the title: as we attempt to imagine this brain that got language, while wondering *why* it got language. What is missing is related to the notion of the intense motivation in the infant to interact with the other human beings around it. Finlay suggests that it is the anatomical link between reward- and motivation-related neural mechanisms in the ventral striatum and prefrontal and orbitofrontal cortex, related to complex behavior, which contributes to the increased responsiveness to socially defined reward and increased disposition to social relationships and attachments (Syal and Finlay 2011).

Arbib's approach is based on the notion of computational neuroethology. Neuroethology is the exploration of animal behavior and of the underlying neurobiology that gives rise to that behavior. Computational neuroethology then is the building of mathematical models that allow one to characterize and study the computations that underlie the behavior. The approach was developed by Arbib already in the Rana Computatrix, whereby a given behavior (in this case visuospatial behavior) is identified, and then the neuroanatomical underpinnings are characterized and modeled. When the same approach is applied to language processing, the unit of behavior studied can reasonably be the sentence. Thus, the sentence is considered in terms of its propositional content, leading to an analysis in the context of grammatical constructions that Arbib and colleagues develop in the format of template construction grammar.

What is missing from this analysis can be found in the motivation to communicate. I will attempt now to argue that language is but the tip of an iceberg whose hidden mass includes a complex infrastructure that must be understood if we are to understand language.

We can explain all of the machinery that allows a system to observe a scene and then produce the sentence “John put the ball on the table” (or to perform the inverse operation) without any requirement on why the system would be motivated to produce such an utterance. Likewise, we could characterize this system in the purely action based framework, where the system can translate between language and images or actions, but such a system does not necessarily have the ability to represent and modify the mental state of the interlocutor, which seems to be such a central aspect of human communication (Grice 1975, Levinson 2006, Tomasello et al. 2005).

Language is a vehicle of human communication, and what makes human communication unique is its social nature, i.e. that it is used by one agent to influence the mental states of another agent. In this context of self and other, the notion of self takes on a central importance. Human language cannot originate without a self.

2.1 Requirement for a self

Arbib’s notion of self as a collection of social schemas (p. 18) can be considered in comparison with notions of self in the developmental psychology domain, for example as proposed by Neisser (Neisser 1997) and his four developmental stages of self. The “ecological self” is the individual situated in and acting on the immediate environment, including the development of the body schema. Infants perceive themselves to be ecological selves from a very early age. The “interpersonal self” is the individual engaged in social interaction with others. This includes communication via gaze, voice, and is considered to become available early in infancy. The “conceptual self” or self concept is considered to involve mental representation largely acquired via verbal information beginning in the second year. Finally, the “temporally extended self” corresponds to the individual’s own lifestory, as one knows, remembers tells and projects it into the future. It is considered that this requires the conceptual self, narratively organized episodic memory, and an understanding of continuity of persons over time.

Arbib proposes a consideration of self as an assemblage of schemas (perceptual schemas, motor schemas, coordinated control programs, etc.) through which experience can be organized (p. 18). While he does not make this explicit, there is no reason why schema theory could not accommodate Neisser’s levels of self. But this should be made explicit, because we will see that it is perhaps the influence of the resulting "social schemas" that will be part of the underlying drive to communicate.

On page 23 Arbib discusses the work of Brothers that "stressed the vital role of social interaction in the evolution and function of animal and human brains," and introduces the Fourth Lampost: Social Schemas. While thus acknowledging the central importance of socialization and the social schema for the development of language, Arbib does not take advantage of the potential impact that the motivational component of social interaction can have on making the brain language-ready. Indeed, Arbib's formulation of social schemas as collective patterns of behavior in a society does not appear to include this motivation to share mental states at the individual level.

In the following section in his chapter on the discovery of mirror neurons, Arbib motivates how the discovery of these neurons provided a powerful explanatory tool, but also one which perhaps monopolized his attention, to the detriment of other aspects of the social brain, including the motivational component.

In the section titled "Motivation and Emotion: The motors of behavior" (Ch. 4, p. 98) there is no development of the child's motivation to communicate, to share mental states.

2.2 Self and other

Chapter 5 briefly alludes to "theory of mind" as the ability to understand and interpret others as being like oneself. Arbib thus suggests that we can consider multiple mirror systems, e.g. for hands, faces, and in humans, language. So that when others speak, my language system is active as if I were speaking. Here, on page 144, Arbib comes perhaps as close as he will to linking emotion to the language system, "by assessing how our ability to experience others' emotions affected the way the language-ready brain has evolved." Part of this underlying motivation to communicate is at the heart of the work of Tomasello and his colleagues (Tomasello et al. 2005). These researchers claim that humans have a unique motivation to share psychological states with others and unique forms of cognitive representation for doing so, including powerful capabilities for intention reading.

Arbib poses questions about how the language-ready brain came about from the biological, cultural and developmental perspectives (p. 158). From the biological perspective, it may be that the linking of the emotion system to the communication system was one of the crucial elements.

In the three key hypotheses outlined in Chapter 6 (p. 161), we could thus consider an additional element that can account for these social abilities: the language-ready brain developed a social cognitive system that is capable of attributing complex mental states to others, and a deeply rooted link to motivation systems that created a strong motivation to share these mental states. This is alluded to in property 7 (p. 166). Language readiness is supported by “genetic coding of brain and body and the consequent space of possible social interactions” (p. 167). It is within this consequent space of social interactions that we want to explore.

2.3 Shared intentions

In citing Tomasello (p. 198) to define the cooperative framework, Arbib is in the right ball park, noting for example that humans will point declaratively while chimpanzee do not, and contrasting the human ability to cooperate with its absence in apes. The key point to derive from the research of Tomasello in this context is that humans are innately very powerfully motivated to communicate. It is this motivation to communicate, to share mental states, to engage socially that is of central importance, and this is revealed in Tomasello’s research in children well before they begin to use language.

In Chapter 11 (p. 284) it is clear that the notion that language is there to allow the infant and caregiver to socially engage is missing. The role of the caregiver is specified in terms of narrowing the attentional focus and thus the search space for learning. The caregiver is not there to direct attention, but rather, the child is motivated to engage and share experience with the adult.

> There is this thing that happens with children: If no one is watching them, nothing is really happening to them. It is not some philosophical conundrum like the one about the tree falling in the forest and no one hearing it; that is a puzzler for college freshman. No. If you are very small, you actually understand that there is no point in jumping into the swimming pool unless they see you do it. The child crying, “watch me, watch me,” is not begging for attention, he is pleading for existence itself.
> –M.R. Montgomery, Saying Goodbye: A Memoir for Two Fathers (Cited in Tomasello, Cited in Chapter 22 Understanding the self as a social agent, in Rochat ed. The self in infancy: theory and research.

2.4 Levinson on the interaction engine

This interacting takes place in the context of parity. The notion of parity that Arbib develops in Chapters 6 and 10 is central to the argument developed in this comment. Arbib’s characterization of parity is twofold, with (i) the hearer recognizing the word the speaker produced, and (ii) having sufficient experience related to that of the speaker so that the schema assemblage that is elicited by the recognition more or less matches the speaker's intentions. But this notion of parity is not sufficiently developed.

Part of the language-ready brain's capabilities must have included what Levinson refers to as the Interaction Engine (Levinson 2006). Levinson argues that before language evolved, there was already an interaction capability that allowed individuals to understand communicative actions not just as behaviors, but as intentional communicative actions in the Gricean sense (Grice 1975). This implies a complex ability to understand intended meanings and communicative intentions that go far beyond the recognition of actions and the parity that could be provided by the mirror system. How did the brain get the interaction engine?

2.5 Finlay on the biological underpinnings

Syal and Finlay (Syal and Finlay 2011) suggest that a new linkage that has occurred in human evolution has been made between the neural representation of central caregivers, the motivational systems, and cortical vocal and gestural learning systems. Thus, influencing the desired individual's attention through vocalization, gaze, contact or localization produces the most profound reward.

In agreement with Finlay and Syal, we would then suggest that this link between profound reward and the ability to influence the desired individual will lead to the development of a social communicative intfrastructure that is part of the language-ready brain. It is likely that this link allowed a co-development between mother and child, with pleasure being derived from both.

2.6 Summary of first comment

Human communication is unlike that of other primates for several reasons, one of the most prominent being the use of spoken language. Language is so striking that it overshadows another perhaps more important aspect of human communication: that is our ability to represent others as agents who have intentional communication like us, and thus as agents upon whom we can count to interpret our communicative behavior in a particular way. This goes beyond the notion of parity that is supported by the mirror system. It entails a very non-trivial form of representation of self, and other as like-self, and non-trivial abilities to operate on these representations.

So "how the brain got language" should also include a focus on "how interacting brains got intentional communication". Considering language in an action-oriented framework should be extended to include language in an intentional communication framework. In order to chart out how the brains got intentional communication, a pathway that is partially illuminated by Finlay would hold that linking powerful motivational systems to perceptual and motor systems for interaction including vocal and gestural signaling would produce a drive to interact with others, and give those others a strong importance. This in turn could produce evolutionary pressures to represent the internal structure of others so as to predict and anticipate, thus leading to more elaborate representations of self and other as required for human-like intentional communication, characterized as the interaction engine by Levinson, and revealed in the myriad of human communicative behavior as characterized in the developing child by Tomasello and colleagues.

3 Compositional semantics and meaning

A second issue that Arbib does not face head on is related to the generative capability of human language, and how that might derive from construction grammar.

3.1 Related construction grammar approaches

Chapter 2 provides "Perspectives on Human Language" which essentially takes a cognitive linguistics stance, whereby grammatical structure is related to conceptual representation and its expression. It would be of interest to consider a comparison of the proposed template construction grammar, and the work of Bergen and (Chang 2005) on embodied construction grammar. Similarly the visually grounded construction grammar is related to work in linking event structure with grammatical structure (Dominey and Boucher 2005, Fern et al. 2002, Hinaut and Dominey 2013, Siskind 2001).

The problem remains as to how grammatical constructions as mappings from semantic action templates onto sentence forms can become generative and address truly novel grammatical forms. This is partially addressed in the transition from holophrase to grammatical construction in development. We note that in his discussion (Jackendoff 2003) of lexical storage versus online construction Jackendoff outlines an approach in which the infant initially is "storing everything," and begins to generalize regular patterns "and extract explicit patterns containing typed variables," allowing the system to "go productive," via variable-based structures. While this addresses how variable based grammatical construction templates can be formed, again, the strong notion of compositionality is still missing.

3.2 Compositional syntax derived from compositional semantics

In Chapter 6 Arbib outlines a set of eleven properties that make the use of language possible. Property 6 is humbly titled "Syntax, Semantics and Recursion." Indeed, this property packs in quite a lot of the business that makes up what can be considered the key properties of language. Arbib considers that syntax specifies how to put words together into constituents, and constituents into larger constituents, etc. Compositional semantics is then what allows one to extract meaning based on the syntactic structure, which leads to recursion: "Recursion in syntax is a corollary to the conceptual structure of what we want to communicate." This is well and good, but it does not address the problem of how the conceptual structure became recursive (which may be beyond the scope of the exercise), nor more importantly, how the mapping from recursive compositional semantic representations to novel syntactically well-formed sentences can take place. We have attempted to address this (Dominey 2003), as the notion that conceptual structure may provide the basis for the recursive structure of language was discussed in a comment on Jackendoff (2003).

In developing a language system in which combinatorial structure exists in phonology, semantics and syntax, Jackendoff's position is consistent with a model in which the combinatorial structure of language serves the purpose of transmitting messages constructed from an equally combinatorial system of thoughts. In this case, the precedence for combinatoriality appears to lie in the thought or conceptual system. The burden of combinatorics is then seen to be pushed onto the conceptual system.

3.3 Neurophysiology of meaning

The question then arises what is the neurophysiological underpinning of meaning? It is of interest in this context that language comprehension appears to recruit the sensory-motor system, as if understanding something is closely related to doing it. This framework is currently often referred to as "embodied language processing," which emerged as an alternative to a much more symbolic approach to language understanding (for review see (Barsalou 1999)).

From the "language-ready brain" perspective, one should be able to respond to the question: how does the anatomical connectivity that supports the embodied meaning system come to be? Looking to current human neuroscience, the semantic/conceptual system (Binder and Desai 2011) has been characterized as a distributed network that includes potential hubs in the frontal, temporal and parietal cortices. These hubs might integrate and systematize the modality-specific semantic features of word meanings that are anatomically distributed across high-level sensory and motor areas of the brain. The question remains, how could these semantic hubs be integrated within the language system? Interestingly there are major white matter pathways linking the perisylvian language areas to these semantic network hubs. A principal connection is provided by major white matter pathways (the arcuate fasciculus and superior longitudinal fasciculus) which link the pars opercularis of Broca's area with posterior temporal and temporo-parietal cortex. An indirect pathway links this perisylvian cortex to parietal cortex, with a second branch descending from this parietal cortex to temporal and temporo-parietal areas (Catani et al. 2005). These direct and indirect pathways provide ample connectivity between Broca's region (perisylvian language areas) and the parietal mirror system.

From a comparative anatomy perspective, the vestiges of the superior longitudinal fasciculus is present in the macaque and chimpanzee, but it is most developed in man (Rilling et al. 2011; Rilling et al. 2008). Thus, we can ask whether the comparative neuroethology approach should take into account such evolutionary changes in the white matter connectivity that correspond to the neuroanatomical organization of the language-ready brain?

3.4 Summary of second comment

The complexity and generativity of the human conceptual system is likely to be equal or superior to that of the language system. It is likely that it evolved in the support of deep representations of the self, the linking of these representations to others in intention based interaction, likely providing the basis for the first expression of generativity in the cognitive domain. The question thus remains, how did the brain get a complex representational system as part of its language readiness?

4 Conclusion

Understanding the origins of language is a treacherous undertaking. As stated above, from the neuroethology perspective, one looks to the behavior in question and its underlying neurophysiology. The tricky issue with language is that of characterizing the behavior. Language can be characterized in one manner as a communicative mechanism that allows the speaker to linearize the contents of a semantic representation so that the listener can reconstruct the intended (potentially multidimensional) meaning (Dominey and Boucher 2005, Dominey et al. 2009). Arbib takes a similar approach. Alternatively, language can be characterized as being but the tip of an iceberg, with the submerged mass being made up of an elaborate interaction engine (Levinson 2006). This requires distinct levels of abilities including the attribution of intention or "mind reading" which corresponds to the mirror system's parity mechanisms. It also requires a second level ability to do the mental computations that allow us to simulate the other simulating us, which is necessary for cooperative interaction underlying communication. A third level corresponds to having Gricean intentions, that is, intentions that drive behaviors whose function is to influence communication by having the underlying intention recognized. Levinson claims that it is this third level that makes high-level communication possible and that laid the foundations for the evolution of language. Arbib has set the groundwork for the inclusion of this analysis in the characterization of the language-ready brain, but work remains to be done.

This three-leveled architecture has been characterized in a different context by Tomasello and his colleagues in their comparative ethology and human developmental research. This research has revealed that humans display a unique motivation to share mental states that is present before the emergence of language, and which can provide further lampposts. Looking at their developmental work, it is striking that prelinguistic (around 18 month-old) infants display an uncanny motivation to link up with the intentional mental states of their adult caregivers. This is expressed over and over by their motivation to show the caregiver new objects, and to help the caregiver complete a failed action. The question arises, what is the origin of this motivation that appears to play such an important role. Finlay and colleagues suggest that we should look to the ventral striatum and its strong anatomical link to communicative systems in order to understand the built in emotional drives that we have to interact with others.

Finally, underlying these interactions is the complex human representational semantic system that has also apparently significantly evolved. The comparative neuroethology approach should not only address the mirror system but also the semantic system, with a concrete example in the comparative neuroanatomy of the white matter pathways which in man links the perisylvian language areas with temporoparietal areas involved in aspects of meaning represention.

Arbib has set out a foundation for the pursuit of these questions.

Acknowledgment: This work is supported by the French ANR "Comprendre" and the EU FP7 Cognitive Systems and Robotics project EFAA.

References

Barsalou L. W. 1999. Perceptual symbol systems. *The Behavioral and Brain Sciences* 22. 577–609.

Bergen, B. & N. Chang. 2005. Embodied construction grammar in simulation-based language understanding. *Construction grammars: Cognitive grounding and theoretical extensions*, 147–190.

Binder J. R. & R. H. Desai. 2011. The neurobiology of semantic memory. *Trends in Cognitive Sciences* 15. 527–536.

Catani M., D. K. Jones & D. H. ffytche. 2005. Perisylvian language networks of the human brain. *Annals of Neurology* 57. 8–16.

Dominey P. & J. Boucher. 2005. Learning to talk about events from narrated video in a construction grammar framework. *Artificial Intelligence* 167. 31–61.

Dominey P., A. Mallet & E. Yoshida. 2009. Real-Time spoken-language programming for cooperative interaction with a humanoid apprentice. *International Journal of Humanoids Robotics* 6. 147–171.

Dominey P. 2003. A conceptuocentric shift in the characterization of language. *Behavioral and Brain Sciences* 26. 674–675.

Fern A., R. Givan & J. M. Siskind. 2002. Specific-to-general learning for temporal events with application to learning event definitions from video. *Artificial Intelligence Research* 17. 379–449.

Grice H. P. 1975. Logic and conversation. In P. Cole & J. Morgan (eds.), *Syntax and semantics*, 41–58. Indiana University Press.

Hinaut X. & P. F. Dominey. 2013. Real-time parallel processing of grammatical structure in the fronto-striatal system: A recurrent network simulation study using reservoir computing. *PloS one* 8(2): e52946. doi:10.1371/journal.pone.0052946.

Jackendoff R. 2003. Précis of *Foundations of language: Brain, meaning, grammar, evolution*. *The Behavioral and Brain Sciences* 26. 651–65.

Levinson S. C. 2006. On the human "Interaction Engine". In S. C. Levinson & N. J. Enfield (eds.), *Roots of Human Society: Culture, Cognition, Interaction*, 39–69. New York: Berg.

Neisser U. 1997. The roots of self-knowledge: Perceiving self, it, and thou. *Annals of the New York Academy of Sciences* 818. 18–33.

Rilling J. K., M. F. Glasser, S. Jbabdi, J. Andersson & T. M. Preuss. 2011. Continuity, divergence, and the evolution of brain language pathways. *Frontiers in Evolutionary Neuroscience* 3. 11.

Rilling J. K., M. F. Glasser, T. M. Preuss, X. Ma & T. Zhao. 2008. The evolution of the arcuate fasciculus revealed with comparative DTI. *Nature Neuroscience* 11. 426–428.

Siskind J. 2001. Grounding the lexical semantics of verbs in visual perception using force dynamics and event logic. *Journal of Artificial Intelligence Research* 15. 31–90.
Syal S. & B. L. Finlay. 2011. Thinking outside the cortex: Social motivation in the evolution and development of language. *Developmental Science* 14. 417–430.
Tomasello M., M. Carpenter, J. Call, T. Behne & H. Moll. 2005. Understanding and sharing intentions: The origins of cultural cognition. *Behavioral and Brain Sciences* 28. 675–691.