Language's matrix

Adam Kendon

In *The origins of human communication,* a book based on the Jean Nicod Lectures delivered in Paris in the Spring of 2006, Michael Tomasello argues that human communication depends upon and is made possible by what he calls a "cognitive infrastructure" which allows humans to share *common ground* with one another. That is to say, humans are able to share in common foci of attention and the interpretative frames in terms of which these foci are understood, and are able each to know this of each other. It is not just that p and q may both attend to the same thing at the same time and in the same way, it is that p knows that q knows to what and how p is attending and that q knows that p knows this, and vice versa. It is this that makes possible the sharing of intentions, the foundation of true cooperation, and this is the foundation of all that is distinctive about human communication. Tomasello argues that *language* is only possible because of this capacity for cooperation that humans have developed. Language, accordingly, is seen as an outcome of this capacity. It is this capacity that makes it possible for actions or vocalisations to come to have meanings in common between individuals, and for it to be possible for shared, conventionalised communication systems to become established. That is, language is an outcome of cognitive and social processes, included within, but not coterminous with, human communication. This is, of course, as Tomasello makes clear (pp. 10-11), the opposite of the position of Chomsky and his followers. Human communication, in its most distinctive character, is the product of biological adaptations for cooperative social interaction in general, while languages, with their lexicons and grammars are "culturally constructed and passed along by individual linguistic communities." As he argues in his penultimate chapter, although there are common principles to the organisation oflanguage, these arise from the general tasks of communication that all languages undertake and not from some inborn system of 'Universal Grammar'. The diversification of lexicons and grammars that specific languages present us with is a product of processes of conventionalisation. 'Language', thus, is not to be understood as the expression of a separate module or organ of the mind. It is a "form of social action constituted by social conventions for achieving social ends" (pp. 343). It is made possible by and grows out of the broader processes of human communication, as indeed the very title of Tomasello's book implies.

The origins of the "cognitive infrastructure"

Studies of the cognitive capacities of apes, mainly of chimpanzees (appropriate animals for comparison, since they are the living species closest biologically to *Homo sapiens),* show that while these animals are able to understand each other as intentional actors, as they can also understand human beings, and that they are able to understand the goals of others and are able to recognise what it is that another may be looking at or attending to, they do not *share* intentions with another and thus do not enter into interactions in which the parties have intentions in common, at the same time both knowing that this is so. Examples of collaborative hunting in chimpanzees observed in the wild in which several chimpanzees together surround and eventually catch a monkey which is then torn apart and eaten (for a summary see Boesch, 2005), which some have claimed to represent the possibility of cooperation, are better understood, so Tomasello suggests, as the consequence of each individual member of the hunting group assuming a line of activity that, in the given situation, is most opportune for the individual. There is no evidence that the chimpanzees share together the common goal of catching the monkey and act, in relation to one another, in accordance with this shared goal. Laboratory studies in which chimpanzees are put in situations in which they are able to obtain something they want only if they act jointly (such as pulling together, but on separate ropes, to obtain food) show that they will do this only in conditions in which they do not have to compete for the food so obtained (Melis, Hare, & Tomasello, 2006 - see p.183).

Humans, on the other hand, readily and typically co-operate with one another. The simplest common exchanges of everyday life entail this. But when does the capacity for this develop? Tomasello describes a series of recent studies in communication between adults and infants between 12 and 18 months of age, concentrating mainly on how and when they point and what pointing is used for. These studies show that infants at this age, before they have developed a capacity to use language, nevertheless are able to use pointing not only to indicate something they would like to have for themselves, but also to indicate things that interest them and that they would like another to take an interest in at the same time, so that their interest can be shared. Further, however, they are able to point to inform another of something, to give them information which may be useful for them. This is a clear indication that the child, even at this age, is able to appreciate the needs and goals of another and to act to help them meet them. Studies are also described which show that infants at this young age, again before having any command of language, are able to share common interpretative frameworks with others. In one study Tomasello cites (Liebal et al., 2009), an adult and child worked together picking up toys and putting them in the toy box. Within this dyad, if the adult pointed to a

toy that was still not in the box, the child responded by taking the toy and putting it away. However, if an adult who had not been involved in this task entered the room and drew attention to toys still out of the box, the child interpreted this in a different way, thus showing that it recognised with whom it shared an interpretative framework.

From observations of this sort, Tomasello concludes that the cognitive infrastructure necessary for cooperative communication is already in place in humans at a very early age, before the child has begun to use language. That is to say, language is *not* the means by which human cooperative communication is made possible as some have supposed (for example, Bickerton, 2005). Rather, it is the presence of this infrastructure that makes language possible. Accordingly, an important part of any inquiry into the phylogenetic origins oflanguage must include an inquiry into the origins of this cognitive infrastructure.

Tomasello proposes that a first step towards this development was the development of a tendency for individuals to be more tolerant of each other and less competitive, presumably in contexts concerning the availability of food. It is not clear what might have led to this development, however. As mentioned earlier, chimpanzees can be persuaded to act in coordination with one another on a task, where the outcome is that of obtaining food sufficient for all parties. In these circumstances, it is perfectly all right to be tolerant. If such an attitude of tolerance can be established as a matter of course, the stage may be set for the next development. That is, in group hunting, for instance, if the catch is always shared (which among chimpanzees it is not) it is no longer so important which individual makes the capture. This makes it possible for others to do things that might be to the direct advantage of the individual most likely to achieve the actual capture, but still of indirect advantage to them, since eventually they will share in the outcome, if it is successful. This could set the stage for groups of individuals coming to act jointly around the common aim of capturing prey. Rather than each trying to do it for himself, they join together to ensure prey capture because, afterwards, all can be assured of a portion. This leads, then to a situation of *mutualism* - where an individual offers help to another because this creates a situation in which he will also benefit.

The next step would be for such helping in immediate mutualistic situations (such as during a hunt) to become generalised to helping more generally. This could come about if individuals gained reputations as helpful individuals in mutualistic contexts. Tomasello mentions a study which suggests that where chimpanzees have a choice of partner in a collaborative task they quickly learn which one is more likely to collaborate and not end up by being aggressive. This could mean that there could be circumstances in which being helpful to others brings valuable rewards to the individual, rewards in the form of being sought out by others as

partners in other contexts. Being helpful to others, thus, implies that others will on other occasions to be helpful to oneself - the principle of "what goes around comes around'; as it were.

Joint orientation and the referential significance of action

Cooperative interaction depends upon participants being able to share objects of attention in common and, furthermore, to be able to share in common the same interpretative perspective with regard to those objects. For this to be possible, however, participants have to be able to let each other know what their object of attention is and that they know what the object of attention of the other is. When an individual *points* for another, he may be said to be *referring* to something, and in many contexts, if A points to X and B also attends to X and does something in relation to it, from the way A may respond to B's attending to and dealing with X, B can learn that the object referred to by Xs pointing is the object that is relevant for A. In this way, both A and B can confirm to one another they are both attending to the same thing, so attention can now be shared. However, although, through joint orientation to something, which may be expressed by each pointing to the same thing, participants can arrive at a situation in which each is simultaneously attending to the same thing, as just described, this process will be greatly facilitated if the object that is the focus of shared attention can be represented or characterised in some way by means of an action that *refers* to it or *stands for* it. When this becomes possible, then we truly are at the threshold of language. The feature of language as a communication system that sets it off from all others is the capacity it affords for someone to refer to things, to represent things, to name them, for others. Tomasello supposes that this would be done at first with a kind of pantomiming. However, as he also points out, for pantomiming to be possible, this not only requires a capacity for mimesis or imitation, it also requires the ability to *recognise* that a mimetic act by another is such an act, and that in engaging in it the actor is *referring* to something. For communication via pantomime to be successful, conspecifics must be able recognise that pantomimes stand for something other than the act itself and they have to recognise that the pantomime itself is an intentional *communicative* act, addressed to another. As Tomasello puts it (p. 203), to comprehend an iconic action as a communicative gesture, one must first "understand ... the Gricean communicative intention; otherwise the recipient will suppose that the communicator is simply acting bizarrely ... "

This is a fundamental question. How does iconic representation come to be possible and how does a recipient come to perceive an iconic gesture as a communicatively intended action, rather than something irrelevant or bizarre? Tomasello

does not inquire into this, but this is something that badly needs to be explored. To do this it would be useful to examine work on the perception of action and how perceivers are able to distinguish intentional action from action that has no purpose and to further explore such work as there is on the perception of action as communicative. From such work as has been done on this, it seems that action recognition is an ability widely shared and has a very deep evolutionary history (see Blakemore & Decety, 2001, for a recent review of human studies and Byrne, 2006, for a discussion of 'behaviour parsing' in primates). The ability to parse the actions of another and to anticipate action outcomes is fundamental if any sort of coordinated social life is to be possible. It is here that a lot more study might pay big dividends for our ultimate understanding how actions come to be understood as 'symbolic' or representational.

From what we have said so far, it already will be clear that such an ability to parse each other's actions is central to the process by which collaborative action becomes possible. As Tomasello argues (quite convincingly, in my view), the establishment of collaborative interaction, can only be achieved by animals mutually perceiving each other's *visible actions*. It could not be achieved by way of vocalisations. A vocalisation can draw the attention of a recipient to its source, to the originator of the sound. But that sound cannot be used by itself to direct the attention of a recipient to something else that is to constitute a specific object for joint attention. One cannot *point* with a vocalisation. This can only be done through a *visible action* that serves to link, in someway, the actor to something in the environment in relation to which he is acting.

When a primate orients its gaze toward something, it is possible for others to see this and, at least in monkeys and apes (as well as in humans, of course), the direction of gaze can be followed, leading another to focus their attention on what it is the looker is attending to. In humans this orientation may also be *expressed* by performing an additional bodily 'movement toward' the focus of attention (see Eco, 1976, p.119), an action that, as one might put it, can be understood as 'showing doing focused looking'. This is generally called 'pointing' which, although commonly done by extending the arm and index finger in the 'movement toward', it can also be done in many other ways as well, as de Iorio ([1832] 2000) described a good many years ago (see also papers by Sherzer, 1972, and Enfield, 2001, on lip-pointing among the Kuna of Panama and among the Lao, respectively, and by Kendon and Versante and Wilkins in Kita, 2003, on the different ways in which the hand may be used in pointing). However it is done, this is the only way in which a focus of attention of one individual can be made visible for another. It is only through the 'reading' of each other's orientations, whether 'mere' orientation, or with the added 'show' of pointing, thus, that *joint orientation* can come about.

This is reminiscent of the process of "frame attunement" (Kendon, 1990, ch. 8) which draws upon Goffman's formulation that when people engage together in what he calls "focused interaction" where, as he has put it, two or a few come to share a common perspective on a spate of events and jointly agree about what is relevant for them together, they effectively enter together into a common attentional frame (Goffman, 1961, 1963, 1974). There are various ways in which participants in interaction can make manifest to one another that they are sharing a common perspective. Here, an important role is played by the jointly sustained systems of spatial-orientational arrangements that are so commonly observed in human occasions of 'jointly focused' interaction (see also Kendon, 1992). These orientational and postural aspects of interactional behaviour, and how these may be temporally synchronised, often in very subtle ways, allow participants to reciprocally inform each other that they are 'together', that they share feeling and attitude with respect to something relevant for them.

Sensitivity to orientation and gaze direction is highly developed among humans, and plays an extremely important role in the process by which they are able to delicately adjust their streams of action to one another within occasions of interaction, as much in occasions of interaction where "talk" is not the "featured mode" of communication (for example, playing tennis or ballroom dancing) as in those where it is. Indeed humans have evolved physical characteristics that playa role in this and which facilitate it. Human eyes only (and not the eyes of any other primates) have white sclera, which means that accurate observation of gaze direction and changes of gaze direction, even over considerable distances, is greatly enhanced (Kobayashi & Koshima, 2001). The complexity and subtlety of facial movements also play an important role. This allows individuals to provide each other with exceedingly delicate information about how they are reacting to one another in a continuous manner (for an example, see Kendon, 1990, Ch. 5). Humans thus have physical and physiological adaptations that make possible very complex and subtle interactions.

These features, which are far more elaborated in humans than in any other social animal, play an indispensable role in co-present interaction and allow for the continuous, detailed and endlessly subtle manner in which people can adjust to one another in interaction. As Tomasello points out, "the linguistic 'code' rests on a nonlinguistic infrastructure of intentional understanding and common conceptual ground, which is in fact logically primary" (p. 58). And it is just these features that we have mentioned, mediated by an anatomical apparatus that makes possible an extreme rapidity and subtlety in co-present interaction, that characterise human interaction.

The primacy of the gestural modality

Tomasello's argument that joint attention can only be established through the co-ordination of visible behaviour is compelling, but creatures that can share foci of attention will be said to have 'language' only when they are able to inform each other as to what it is that each is attending in common. As we have already discussed, I may point at something and so refer to it, and you may point at it, likewise. However, unless there is a way by which we mutually inform each other of what it is we are both attending to, we do not have shared attention. It is at this point, it seems, that an ability to name things, to refer to them in a symbolic way, thus becomes important. The question is, how might this be done? As we have already mentioned, Tomasello suggests that this might have been done at first by means of pantomimic or iconic gestures which, he thinks, could be derived originally from the ritualised intention movements that are the basis for gestures in apes. He believes that 'pantomime' or 'iconic gestures' would have been the first means by which a vocabulary of shared expressions would have been developed, constituting, thus, the first form of language.

He puts forward four arguments in favour of this idea. These will be familiar to those who have followed recent discussions of this view of language origins, as presented in the work of Hewes (1973), Corballis (2000), or Armstrong and Wilcox (2007). These include the point that ape vocalisations are too fixed in form, too tied to emotion and insufficiently under voluntary control to offer a model for something that can be imagined as a precursor to speech; the gestural expressions of apes, on the other hand, are highly flexible and modifiable by learning and thus seem to provide a basis from which voluntary gestural expressions can readily be fashioned; children use gestures, both pointing and pantomimic expressions, before they begin to speak; and finally, in circumstances where, for various reasons, speech cannot be used (including deafness, speech-avoidance taboos, or high levels of persistent noise), humans are able to create very rapidly complex linguistic systems using gestures alone. For Tomasello, such "gestural inventions" would be "incredible" if we did not suppose that humans were adapted to gestural communication first (p. 328).

The problem of speech

As even the proponents of the 'gesture first' view agree, however, and as opponents of this view are always eager to remind us, modern languages are overwhelmingly *spoken*. Although the complex gesture systems developed in deaf communities can appropriately be referred to as languages and are regarded, in principal, as being

no different in their 'languageness' from German or Quechua, most would agree that these are adaptations to special circumstances and do not challenge in any way the unavoidable observation that humans are *speaking* animals. They may be testimony to the idea that the faculty for 'language' is not inevitably tied to a specific modality, but they do not lead to a revision of the view that *speech* is the vehicle for language which humans are biologically adapted to use. Anatomical, neurological and physiological adaptations to the process of *speaking* are something that everyone is born with, and all baby humans (even deaf ones) enter a phase in which they begin to exercise their speaking apparatus by 'babbling'. As Lenneberg (1967) made clear many years ago, humans are biologically predisposed to *speak,* even if, for some, speaking is not, or cannot be, the main vehicle for language. This being so, a great difficulty is created for 'gesture firsters'

Tomasello discusses this problem in a section entitled "The switch to the vocal modality" (pp. 226 ff.) and he expresses the view that vocal language "is actually a recent overlay" (p. 246). Like Michael Corballis, he seems to be believe that humans *first* developed language in gesture and then, later, changed over to using speech. To account for why this happened he invokes the various advantages of spoken communication that have so often been proposed, such as the fact that it frees the hands for other activities while one is speaking, it enables communication over longer distances or in dense forests or in the dark, and so on. To these he adds the idea that vocal language is advantageous because it is more unavoidably public than gesture, and thus may play a role in creating social reputations. He thinks that the earliest vocal utterances were emotional accompaniments for already meaningful action-based gestures, or perhaps "added sound effects" to them (p. 231). Then, "[a]s humans gained more voluntary control over their vocalisations, they could have also used some vocal icons (e.g. making sounds of a leopard) ... [b]ut at some point, in some situations, the vocalisation came to be functional on its own – perhaps under pressure to communicate at long distances, or for communication to be in the public space, and so forth" (p. 232).

These suggestions, like those of most other writers on this topic, are not very helpful. The evolutionary emergence of speech cannot be explained by referring to the apparent advantages that it now might have over other forms of communication. Further, Tomasello offers us no hypothesis at all as to how or why humans "gained more voluntary control" over their vocalisations. Any theory about the origin of language (whether or not it is a "gesture first" theory) must have something to say about this, but on this issue Tomasello remains silent (like almost all other writers on this topic, I might add).

Difficulties and a possible path to a solution

I think much of the difficulty that Tomasello and other promoters of the 'gesture first' theory of language origins have with accounting for spoken language stems precisely from the idea that Tomasello himself expresses, that somehow there must have been a "switch" from gesture to speech at some point. This idea of a "switch" follows, if one thinks that there first developed a language wholly in gestures, which was then later replaced by a language wholly spoken – "Conventional languages (first signed and then vocal)", as Tomasello says on p. 9. However, as MacNeilage (1998) has asked, if an adequate language was first developed in the gestural modality, why was it ever given up? The advantages claimed for spoken language cannot provide an answer to this question that is consistent with modern evolutionary thinking. Speech was not invented to overcome the disadvantages of signing. Extensive communicative uses of vocalisation must have already been developing over a very long period, if extensions of its use for language were to occur.

A much better approach, it seems to me, and one that takes into consideration how utterances are actually produced in modern speakers, would be to start with the assumption that the transition into referential or language-like expressions involved hands and body, face and voice and mouth, all together, as an integrated ensemble. What so many writers on this topic – "gesture firsters" and "speech firsters" both – pay little attention to is the fact that modern humans, when they speak together in face-to-face situations, especially in the informal settings of everyday interaction, *always* mobilise face and hands and voice together in complex orchestrations. Indeed, kinesic and audible expressions are so intimately intertwined in their employment in the creation of utterances, they must be seen as being produced under the guidance of a single project (Kendon, 1972, 1980, 2004; McNeill, 1992, 2005). Verbal language is *never* spoken as words only. It is true that we can write down people's words and we can often present their utterances as something that seems to be made up only of words, but this is only an abstraction from what people actually do when they construct utterances. Every single utterance using speech employs, in a completely integrated fashion, patterns of voicing and intonation, pausings and rhythmicities, which are manifested not only audibly, but kinesic ally as well, and always, as a part of this, there are movements of the eyes, the eyelids, the eyebrows, the brows, as well as the mouth, and patterns of action by the head. And there are, in addition, from time to time, variously conspicuous hand and forearm actions or 'gestures' (as they are called – and which tend to receive the lion's share of academic attention), and also postural and orientational changes. All of these are produced in a fully integrated fashion and must be seen as inseparable components of the utterance as the utterer produces it.

This is very well known, of course, and in modern times it has been repeatedly recognised in academic publications, at least since the work of Birdwhistell (1952; see also Birdwhistell, 1970). Its implications for theories oflanguage and its evolution are rarely appreciated, however. This is because, almost always, 'language' is thought of as a self-contained, autonomous system that is confined to only one modality, and it is the origin and evolution of language thought of in this way that typically preoccupies students of glottogenesis. For a long time, the only modality that language was thought to have was 'speech' (written language was a mere derivative form, of little linguistic interest), but since the recognition of the linguistic status of sign languages (beginning in the 1970s), it became generally agreed that one can have a language in either the spoken or the kinesic modality. When this happened, however, the concept of 'language' as a self-contained system was simply extended to include sign languages. As a result they, like spoken languages, came to be regarded as well demarcated fully autonomous systems. Sign languages may be organised within the kinesic modality, nevertheless, they are to be kept rigourously separate from any uses of the kinesic modality that might be found in speakers. In the case of sign languages, this has led to discussions about how to separate 'gesture' from 'sign' and, recently, to the issue of whether signers, like speakers, also use gestures (for an historically informed discussion of this, see Kendon, 2008).

The mono-modalic view of language, I believe, has become dominant because of the technology of writing. As has been pointed out by several scholars in recent years, including Harris (1981) and Linell (2005), the 'language' of linguistic study is what can be written down. The consequence of this has been that those aspects of the activity of utterance production that cannot be accommodated by the technology of writing tend to be treated by students of language as in some way peripheral. The communicative importance of these 'paralinguistic' aspects is often recognised, but generally they are not thought of as 'language' in *sensu strictu,* and for this reason they have not usually been a part of the 'target' of language origins discussion.

When it comes to sign languages, a rather similar model has been applied, but because no written forms of these languages exist, the problem of keeping 'paralinguistic' features separate from 'linguistic' features is much more difficult. In recent years it has become clear that central to the construction of utterances in sign languages are forms of expression such as 'classifiers', 'constructed action; or 'highly iconic forms' (on this, see Cuxac and Sallandre, 2007), as well as an exploitation of space that is not possible in speech, all of which look very similar to various kinesic devices used by speakers (albeit apparently not in as systematic a manner). This leads to the suggestion that when comparing spoken and signed language, the comparison should be with language as it is performed in speaking, for it is only the performed version of a sign language that is ever available. This

raises the question as to whether, after all, from the point of view of how utterances are constructed, it is as essential to view what speakers do as an integrated performance, as this has been recognised as necessary for some time, in the study of signed discourse.

Given this, then, the 'natural' state of spoken language is a *speech-kinesis ensemble* and, we must suppose, this has always been the case. With the development of writing, and its ultimate emergence as an autonomous form of language with its own properties, which, nevertheless, has provided the dominant model for what 'language' is, at least since the end of the eighteenth century, we have ceased to see how gestures and other aspects of utterance performance are a part of "what is said:' The recent history of language has favoured the separation of modalities. In many glottogenetic discussions, it is this separated modality of written-down spoken language (which dominates our conception of language) that tends to be projected backwards to the earliest days of language, making it very difficult to imagine how it might have arisen. If the same kind of thinking is applied to how we think of sign language and if we then suppose that sign languages arose first, projecting back into the earliest times something like a modern sign language (see Corballis, 2000, p. 125), it is no wonder that we think there must have been a 'switch', and are much puzzled by this.

In my view, there never was a 'switch'. Forms of communicative action which acquired functions of symbolic reference and thus functioning communicatively in a language-type manner, were 'polymodalic' from the beginning. It is possible, as Tomasello argues, that the very first kinds of actions that served as 'representations' or 'symbolic expressions' were visible actions, that is 'gestures' ('pantomimes' and 'pointings', as he supposes), and not actions involving the voice, but 'speaking' must have already been available even in those distant times. We may envisage a progressive incorporation of speaking into the linguistic function, but there never was the 'switch' that puzzles Tomasello, which bothers MacNeilage, and which also has challenged Corballis and, before him, Hewes, as well as many other commentators.

How did this progressive incorporation of the vocal modality into linguistic functions come about? In thinking about this it it helpful to consider the relationship between the control systems involved in the oral movements of speech and those involved in the control of the hands and arms. There is considerable evidence from neurological and other kinds of studies that suggests that there is an intimate and intricate overlap between the neurological apparatus that controls manual actions and the neurological apparatus that controls the actions employed in speaking. This suggests that the relationship between mouth actions and hand actions have a very deep phylogenetic origin. As Gentilucci and Corballis (2006) have proposed, this may have had its origins in the development of hand-mouth coordination that came to be increasingly important as primate ancestors

developed ever more complex and flexible strategies for living in a three dimensional arboreal environment. In such an environment forelimbs came to be used for grasping food and transporting it to the mouth. This could have provided the origin of the nexus of control systems that in modern humans makes possible the co-involvement of hand and mouth in the activities of using language.

Speech itself, when viewed as a complex activity involving the voluntary production of actions of the jaws, lips, and tongue (combined with vocalisation) could have arisen, as MacNeilage (2008) has suggested, as an elaboration of the manipulatory activities of these organs, as they are involved in the seizing and mastication of food. We may suppose that these movements, from their beginning, would have been coordinated with hand actions of grasping, pulling and carrying food to the mouth. It was this ensemble of mouth and hand movements that was exapted for what would eventually become symbolic or referential communication, and provided the foundations for the eventual elaboration of such communication into what is today referred to as "language". That is to say, in the line leading to *Homo,* hand actions and mouth actions have always been in coordination. The hand actions that are coordinated with speaking, as we see them today, did not evolve so as to enable a mimetic form of communication that is necessary for the expression of things that are otherwise difficult for speech to manage (see Goldin-Meadow and McNeill 1999), nor are they "left over" from an earlier stage in which language was gestured rather than spoken (rather as Hewes, 1973, seems to have supposed). Rather, the co-involvement of hand actions with mouth actions in speech is an exaption of the whole complex of actions that are involved when the individual engages in acting on the environment, especially, perhaps, in feeding, but in all other ways as well. Hand and mouth coordination was there from the beginning as part of what we might call an ensemble of *praxic* activity. It was through a transformation of praxic activity into 'virtual' or 'as if' modes that made possible the emergence of action as 'representation', so making the emergence oflanguage possible (Kendon, 1991). It is because this ensemble was developed as a hand mouth coordination system that we can explain why manual actions are co-involved in utterance productions in modern humans.

The origins of grammar and language diversification

Tomasello's account, up to this point, shows that the establishment of common ground and the ability for individuals to share intentions together is crucial for the development of referentially significant actions, making possible the emergence of forms of action that are mutually understood as making reference to things. This is what we have to have if we are to have anything we might want to refer to

as a 'language'. However, in language, of course, far more is done than the establishment of common perspective and the reciprocal informing by participants of what it is they are attending to. Language is a system that allows us to elaborate our requests to one another, to make statements about the state of things in the world and to give accounts of what we ourselves and others have done and thought and many other things. Furthermore, despite certain universal principles by which languages seem to operate, the specific systems in use are quite diverse. As Tomasello says "what we need is an account that will enable us, in the end, to explain the emergence *not* of Language (with a capital L), but the emergence of 6000 different human languages with 6000 different sets of communicative conventions ... " (pp. 243-244). Accordingly, in his penultimate chapter he undertakes to sketch in some ideas about how grammatical complexity came about, and also how linguistic change and diversity comes about.

In his exposition of the characteristics of human cooperative communication (in Chapter 3), Tomasello introduces the idea that there are three fundamental kinds of motivation that give rise to it. These are *requesting* (as when I draw your attention to something because I want it, and you can help me by getting it for me); *offering help* or *informing* (in which I draw your attention to something because you knowing about it will be helpful to you); and *sharing* (in which I draw your attention to something and, as I do so, express my interest in it and attitude toward it, because I want you to orient to that thing in the same way. For example, when I say excitedly, "oh look at that rainbow!" because I want you to be excited *with me* about that rainbow). According to Tomasello these three kinds of motivation are "species unique" for humans and they underlie all of the communicative acts that we engage in.

Tomasello then suggests (in Chapter 6) how these three fundamental motivations can account for different degrees or levels of grammatical complexity that are found in languages. He suggests that, according to the purpose one may have for producing a communicative act for another, the amount and kind of information that one needs to provide in such an act will vary. If all I want to do is to request something, only myself and you are involved, no great syntactic complexity need be involved when I formulate my request. However, if my motive is to let you know about something this may mean that events and actors remote in time and space may have to be referred to, and to do this, using a vocabulary of significant action units ('words' or 'signs') may require the use of syntactic marking devices of considerable complexity. If, further, I wish not merely to inform but to induce you to share feelings and attitudes, "if I wish to share ... in the narrative mode ... a complex series of events with multiple participants playing different roles in different events" (p. 245) the syntax I require to do this will be even more complex. Accordingly, Tomasello proposes that the grammar of a language will be elaborated differentially,

according to the communicative tasks that is being made to serve. Rather quaintly, he refers to the grammar of requesting as "simple syntax", the grammar of informing "serious syntax" and the grammar of sharing as "fancy syntax".

Consistent with this view, Tomasello argues that there are broad principles that govern language organization. These, however, are not derived from some Chomskyan 'Universal Grammar' – which, in Tomasello's opinion "has no coherent formulation" (p. 312). Rather, as he writes, "there are universals of language because people all over the world have similar communicative jobs to get done and similar cognitive and social tools with which to do them." Despite this, there is great diversity. How is this to be explained? According to Tomasello" [t] he most plausible explanation for this ... is the need for groups of humans to differentiate themselves from other groups" and, at the same time, for a given group the particular language it comes to develop, because of the way in which attitudes and experiences in common can be shared, language peculiarity for a given group "is a major way that cultural groups create their own internal group identities."

These are valid points, yet this sort of explanation seems to suggest that languages diversify because it is socially convenient for them to do so. Diversification arises *because* human groups *need* to differentiate themselves from one another. This is another example of explaining the origin of something in terms of its present day utility, in a vein rather similar to those explanations Tomasello advanced for why there was a "switch" to vocal language. A more plausible approach would be to see that the diversification of languages (and indeed of regional and local varieties within given language areas) comes about as a consequence of the way in which linguistic expressions serve to signify things in the world. They do this, not by signifying things directly but by signifying the *concepts* in terms of which the things in the world are construed. These concepts or conceptual categories are not dictated by anything in the world, they are, rather, the *creations* of communities of speakers. A language such as French, for instance, to an important extent is a socially shared system of conceptual construals with linguistic forms as the exponents of these construals.

Participants coming to establish meanings for significant actions in common do this locally in relation to very local requirements and these are bound to be different from one local community to another. There are no cognitive categories that are 'given' (except possibly at some very abstract level), so that the way one group may end up setting up linguistic categories can be rather different from the way another group may do this. The categorisations that language creates are the products of socially shared agreements, and these are free to vary within a very wide range. Thus we might expect groups not in contact with one another to separately produce different languages, just on these grounds alone. And, by the same token, if a group divides and the parts go their separate ways, over time there will

be linguistic differentiation. Language differences can be (and are) exploited as sources of group pride and identity, and people may work to create and exaggerate such differences as part of the process of group differentiation, but the fundamental reason for language differentiation lies in the the fact that languages are conceptual categorisation systems, freely created through local historical processes.

Conclusion

Tomasello argues that human communication emerged from a mode of interaction in which participants became able to share each other's intentions and this, in turn, arose through a modification in the way in which collaborative interactions were carried out, such that participants assisted each other in working toward a common goal. Fundamental to this development was the development of the capacity for interacting partners to simultaneously attend to the same thing. Reference is first achieved through the "natural" gesture of pointing, but if participants come to be able to *name* for each other the objects of their attention, some kind of symbolic representation is required and this, Tomasello believes, was first attained through what he calls the "natural" gesturing of pantomime.

No hypothesis about the origins of pointing is offered, but it is suggested that pantomiming, or iconic gesture could have derived from manual and other intention movements. In the ways in which apes employ their hands and arms in interaction with each other, especially in play, we can see a range of flexible, learned actions which serve as signals of participants' intentions and we may perhaps imagine how these could become stable in form, through processes of conventionalisation, although how the transition is made to such actions serving referentially or as names for things remains obscure.

Once collaborative interactions are established, the forms of communicative action, mediated by pointing and pantomiming, can become conventionalised, leading to the establishment of shared traditions of communicative practices. This first happened in the gestural modality, the vocal modality coming to be overlaid or "piggybacked" (p. 330) upon this shared system. This can rapidly lead to the emergence of vocal languages. Grammatical structures develop coordinately with the development of the different kinds of communicative tasks that language comes to be used for. Basically these can be analysed as *requesting, informing* and *sharing*. He suggests that for each type of communicative task, different elaborations of syntax will come about. Because we are now dealing with social traditions in the development of language, these traditions diverge, and this is what gives rise to the great diversity of language systems we see today, a diversification which, Tommasello believes, has been driven by a need for human groups to differentiate

themselves from each other and to develop special forms of shared action as a way of enhancing group identity.

In my discussion I have tried to bring out some of the difficulties with this account. One of the most important of these has to do with the problem of how the 'natural' gestures of pointing and, more especially, perhaps, of pantomiming, come to be recognised as such. Tomasello sees this problem, at least with regard to 'pantomime', but offers no solution to it. *How* conspecifics come to recognise that each other's actions can be symbolic or referential is not yet understood. The other important difficulty is that an adequate theory for the origin of speech is not provided. Here we have suggested that some modifications of MacNeilage's theory of speech origins in which the hand and mouth ensemble is viewed as a practical, manipulatory system, might supply a useful starting point. We argue that to understand how language has evolved it is necessary to look at what speakers actually do when they construct utterances: they *always* do this using kinesic and oral resources in a complex orchestration. To continue to consider the system to be explained as the abstracted, autonomous system that writing suggests (and this has been extended to inform to our conception of sign languages in a similarly misleading manner), will only leave us trapped with the problem of how there could have been a 'switch' from gestural language to vocal language. As we have argued, there never was such a switch.

Any attempt to sketch out the origins of language in so broad and comprehensive a manner as is attempted in this book is bound to have shortcomings. Readers should not be deterred, however. What is especially attractive about what has been done here is the way in which we are shown the nature of the cognitive and social *matrix* which language surely required. Tomasello succeeds in focusing our attention on some quite specific issues which need to be further explored if our understanding of the problem of language origins is to be advanced. It is the great merit of his exposition that he is able to show us how the elements of this matrix must fit together and shows how experimental work with both non-human primates and human infants can be truly useful for this problem. Above all, it is the overall frame within which the problem of language origins is set, that makes this book well worth serious consideration.

Note

* Michael Tomasello (2008). *The origins of human communication*. Cambridge, MA: The MIT Press. A Bradford Book. Pp. xiv+393.

References

Armstrong, David F. & Sherman E. Wilcox (2007). *The gestural origin of language*. Oxford: Oxford University Press.

Bickerton, Derek (2005). Language first, then shared intentionality, then a beneficient spiral. *Behavioral and Brain Sciences, 28* (5),691-692.

Birdwhistell, Ray L. (1952). *Introduction to kinesics: an annotation system for analysis of body motion and gesture*. Louisville, KY: University of Louisville.

Birdwhistell, Ray L. (1970). *Kinesics and context: essays in body motion communication*. Philadelphia: University of Pennsylvania Press.

Blakemore, Sarah- Jayne & Jean Decety (2001). From the perception of action to the understanding of intention. *Nature Reviews Neuroscience*, 2, 561-567.

Boesch, Christophe (2005). Joint cooperative hunting among wild chimpanzees. *Behavioral and Brain Sciences*, 28, 692-693.

Byrne, Richard W (2006). Parsing behavior: a mundane origin for an extraordinary ability? In N. J. Enfield & Stephen C. Levinson (Eds.), *Roots of human sociality: culture, cognition and interaction* (pp. 478-505). Oxford: Berg.

Cuxac, Christian & Marie-Anne Sallandre (2007). Iconicity and arbitrariness in French Sign Language: highly iconic structures, degenerated iconicity and diagrammatic iconicity. In Elena Pizzuto, Paola Pietandrea, & Raffael Simone (Eds.), *Verbal and signed languages: comparing structures, concepts and methodologies* (pp.13-33). Berlin: Mouton de Gruyter.

Eco, Umberto (1976). *A theory of semiotics*. Bloomington, Indiana: Indiana University Press.

Enfield, N. J. (2001). 'Lip pointing': With a discussion ofform and function with reference to data from Laos. *Gesture*, 1, 185-212.

Goffman, Erving (1961). *Encounters*. Indianapolis: Bobbs-Merrill,

Goffman, Erving (1963). *Behavior in public places*. New York: The Free Press of Glencoe. Goffman, Erving (1974). *Frame analysis*. Cambridge, MA: Harvard University Press. Goldin-Meadow, Susan & David McNeill (1999). The role of gesture and mimetic representation in making language the province of speech. In Michael Corballis & Stephen Lea (Eds.), *The descent of mind* (pp. 155-172). Oxford: Oxford University Press.

Harris, Roy (1980). *The language makers*. Ithaca, NY: Cornell University Press.

Hewes, Gordon W (1973). Primate communication and the gestural origins of language. *Current Anthropology*, 14,5-24.

Iorio, Andrea de (2000 [1832]). *Gesture in Naples and gesture in classical antiquity. A Translation of* La mimic a degli antichi investigata nel gestire napoletano *by Andrea de Iorio* (1832) *and with an introduction and notes by Adam Kendon*. Bloomington: Indiana University Press.

Kendon, Adam (1972). Some relationships between body motion and speech. An analysis of an example. In Aaron Siegman & Benjamin Pope (Eds.), *Studies in dyadic communication* (pp.I77-210). Elmsford, NY: Pergamon Press.

Kendon, Adam (1980). Gesticulation and speech: two aspects of the process of utterance. In Mary Ritchie Key (Ed.), *The relationship of verbal and nonverbal communication* (pp. 207- 227). The Hague: Mouton and Co.

Kendon, Adam (1990). *Conducting interaction: patterns of behavior in focused encounters*. Cambridge: Cambridge University Press.

Kendon, Adam (1991). Some considerations for a theory of language origins. *Man*, (N. S.) 26, 602-619.

Kendon, Adam (2004). *Gesture: visible action as utterance*. Cambridge: Cambridge University Press.

Kendon, Adam (2008a). Some reflections on 'gesture' and 'sign: *Gesture,* 8 (3),348-336.

Kendon, Adam (2008b). Signs for language origins? *The Public Journal of Semiotics,* 2 (2) 2-29

Kendon, Adam & Laura Versante (2001). Pointing by hand in 'Neapolitan'. In Sotaro Kita (Ed.), *Pointing: where language, culture and cognition meet* (pp.109-137). Mahwah, NJ: Lawrence Erlbaum.

Kita, Sotaro (Ed.) (2003). *Pointing: where language, culture and cognition meet*. Mahwah, NJ: Lawrence Erlbaum.

Kobayashi, Hiromi & Shiro Koshima (2001). Unique morphology of the human eye and its adaptive meaning: comparative studies of external morphology of the primate eye. *Journal of Human Evolution,* 40, 419-435.

Lenneberg, Eric H. (1967). *Biological foundations of language*. New York & London: John Wiley and Sons.

Liebal, Kristin, Tanya Behne et al. (2009). Infants use shared experience to interpret pointing gestures. *Develomental Science,* 12 (2),264-271.

Linell, Per (2005). *The written language bias in linguistics: its nature, origins and transformations*. London & New York: Routledge.

MacNeilage, Peter F. (1988). Evolution of the mechanism oflanguage output: comparative neurobiology of vocal and manual communication. In James R. Hurford, M. Studdert - Kennedy, & Chris Knight (Eds.), *Approaches to the evolution of language* (pp.222-240). Cambridge: Cambridge University Press.

MacNeilage, Peter F. (2008). *Origin of speech*. Oxford: Oxford University Press. McNeill,

David (1992). *Hand and mind*. Chicago: University of Chicago Press.

McNeill, David (2005). *Gesture and thought*. Chicago: University of Chicago Press.

Sherzer, Joel (1972). Verbal and nonverbal deixis: the pointed lip gesture among the San Bias Cuna. *Language in Society,* 2 (1), 117-131.

Wilkins, David P. (2003). Why pointing with the index finger is not a universal (in cultural and semiotic terms). In Sotaro Kita (Ed.), *Pointing: where language, culture and cognition meet* (pp.171-215). Mahwah, NJ: Lawrence Erlbaum.