On the Different Origins of Symbols and Grammar

Michael Tomasello

DOI:10.1093/acprof:oso/9780199244843.003.0006

Abstract and Keywords

This chapter emphasises the role of psychology in language evolution, but claims that it was the separate evolution of capacities for using symbols and grammar (that is, syntactic structure) that distinguishes human communication from the communication of other primates. It suggests that there was no specific biological adaptation for linguistic communication. Rather, there was an adaptation for a broader kind of complex social cognition that enabled human culture and, as a special case of that, human symbolic communication. A crucial part of this adaptation was an evolved ability to recognise other individuals as intentional agents whose attention and behaviour could be shared and manipulated. The capacity for grammar subsequently developed, and became refined through processes of grammaticalisation occurring across generations — but with no additional biological adaptations. In support of this perspective, psychological data from the study
of language development in young children and from comparisons with the linguistic, social, and mental capacities of nonhuman primates are presented. More generally, this chapter sees the origin and emergence of language as merely one part in the much larger process of the evolution of human culture.

Keywords: symbols, grammar, language, language evolution, psychology, biological adaptation, social cognition, grammaticalisation, nonhuman primates, culture

Human communication is most clearly distinguished from the communication of other primate species by its use of (1) symbols and (2) grammar. This means that progress on questions of language origins and evolution depends crucially on a proper understanding of these two phenomena. To state my own bias up front, I believe that symbols and grammar need to be investigated from a more explicitly psychological perspective than has been the case in the past. Further, I believe that we will make the most rapid progress if we investigate first and in detail extant systems of communication that are simpler than full-blown human language-specifically, those of human children and those of our nearest primate relative, the chimpanzee. What I would like to do in this chapter, therefore, is to look first at symbols and then at grammar from the point of view of the most basic social-cognitive and communicative processes involved— to see if this can help us in our thinking about how language might have evolved in the human species.

Symbols

This is not the place to review all of the many approaches to linguistic symbols that exist on the current intellectual scene. Suffice it to say that criteria such as arbitrariness and spatial-temporal displacement are decidedly unhelpful when looking at actual communicative processes. After all, Pavlov’s dog associated the arbitrary sound of a bell with food that was not at the time perceptually present. And the expressions ‘duality of patterning’ and ‘stands for’ (as in the locution ‘symbols stand for their referents’) simply put a new name on the phenomenon without providing any further insights.
My proposal is that symbolic communication is the process by which one individual attempts to manipulate the attention of, or to share attention with, another individual. In specifically linguistic communication, as one form of symbolic communication, this attempt quite often involves both (a) reference, or inviting the other to share attention to some outside entity (broadly construed), and (b) predication, or directing the other’s attention to some currently unshared features or aspects of that entity (in the hopes of sharing attention to the new aspect as well). Comprehension of an act of symbolic communication thus consists in understanding that ‘She is attempting to direct my attention to X’ or ‘She is attempting to direct my attention to Y with respect to X’.

We may make this account a bit more concrete by looking at some very young human infants. Six-month-old infants interact dyadically with objects, grasping and manipulating them, and they interact dyadically with other people, expressing emotions back-and-forth in a turn-taking sequence. If people are around when they are manipulating objects, the infants mostly ignore the objects. If objects are around when they are interacting with people, they mostly ignore them. But at around 9-12 months of age a new set of behaviours begins to emerge that are not dyadic, like these early behaviours, but triadic in the sense that they involve infants coordinating their interactions with objects and people, resulting in a referential triangle of child, adult, and the object or event to which they share attention. Most often the term ‘joint attention’ has been used to characterize this whole complex of social skills and interactions (see Moore and Dunham 1995). Most prototypically, it is at this age that infants for the first time begin flexibly and reliably to look where adults are looking (gaze following), to engage with them in relatively extended bouts of social interaction mediated by an object (joint engagement), to use adults as social reference points (social referencing), and to act on objects in the way adults are acting on them (imitative learning). In short, it is at this age that infants for the first time begin to ‘tune in’ to the attention and behaviour of adults on outside entities.
Not unrelatedly, at around this same age infants also begin actively to direct adult attention and behaviour to outside entities using deictic gestures such as pointing or holding up an object to show it to someone. These communicative behaviours represent infants’ attempts to get adults to tune in to their attention and interest to some outside entity. Also important is the fact that among these early deictic gestures are both imperatives, attempts to get the adult to do something with respect to an object or event, and declaratives, attempts to get adults simply to share attention to some object or event. Declaratives are of special importance because they indicate especially clearly that the child does not just want some result to happen, but that she really desires to share attention with an adult as an end in itself. It is thus the contention of some theorists, including myself, that the simple act of pointing to an object or event for the sole purpose of sharing attention with someone else is a uniquely human communicative behaviour, the lack of which is also a major diagnostic for the syndrome of childhood autism (e.g. Tomasello 1995; Gómez et al. 1993; Baron-Cohen 1993).

Carpenter et al. (1998) followed a group of infants longitudinally from 9 to 15 months of age as they engaged in nine different joint attentional activities, everything from following gaze to imitating actions on objects to declarative pointing. They found that for any individual child all nine skills emerged together as a group within just a few months, with some predictable orderings among individual skills and with correlations among the ages of emergence as well. The almost simultaneous ontogenetic emergence of these different joint attentional behaviours, and the fact that they emerge in a correlated fashion, suggests that they are not just isolated cognitive modules or independently learned behavioural sequences. They are all reflections of a single cognitive change: infants’ dawning understanding of other persons as intentional agents. Intentional agents are animate beings who have goals and who make active choices among behavioural means for attaining those goals, including active choices about what to pay attention to in pursuing those goals. This new understanding of other persons represents a veritable
revolution in the way infants relate to their social and cultural worlds.

This social-cognitive revolution at 1 year of age sets the stage for the infants’ second year of life, in which they begin imitatively to learn the use of all kinds of tools, artefacts, and symbols—that is, cultural entities with an intentional dimension, things that point beyond themselves to other outside entities. Thus, tools point to the problems they are designed to solve and linguistic symbols point to the phenomena they are designed to indicate. Therefore, socially to learn the conventional use of a tool or a symbol, children must come to understand why, toward what outside end, the other person is using the tool or symbol, that is to say, the intentional significance of the tool use or symbolic practice—what it is ‘for’, what ‘we’, the users of this tool or symbol, do with it (Tomasello 1998a). For example, in a study by Meltzoff (1988) 14-month-old children observed an adult bend at the waist and touch his/her head to a panel, thus turning on a light. They followed suit, presumably thinking, “This is how this artefact works”. Infants thus engaged in this somewhat unusual and awkward behaviour, even though it would have been easier and more natural for them simply to push the panel with their hand. One interpretation of this behaviour is that infants understood that (a) the adult had the goal of illuminating the light and then chose one means for doing so, from among other possible means, and (b) if they had the same goal they could choose the same means.

Cultural learning of this type thus relies fundamentally on infants’ tendency to identify with adults, and on their ability to distinguish in the actions of others the underlying goal and the different means that might be used to achieve it. This interpretation is supported by the more recent finding of Meltzoff (1995) that 18-month-old children also imitatively learn actions that adults intend to perform, even if they are unsuccessful in doing so. Similarly, Carpenter et al. (1998) found that 16-month-old infants will imitatively learn from a complex behavioural sequence only those behaviours that appear intentional, ignoring those that appear accidental. Young children do not just mimic the limb movements of other
persons, they attempt to reproduce other persons’ intended actions in the world.

The main point in the current context is this. Although it is not obvious at first glance, something like this same imitative learning process must happen if children are to learn the symbolic conventions of their native language. Although it is often assumed that young children acquire language as adults stop what they are doing, hold up objects, and name these objects for them, this is empirically not the case. Linguistics lessons such as these are characteristic of only some parents in some cultures and for only some kinds of words (e.g. no one names for children acts of ‘giving’ or prepositional relationships such as ‘on’ or ‘for’). In general, for the vast majority of their language children must find a way to learn a new word in the ongoing flow of social interaction, sometimes even from speech not addressed to them (Brown 2001). In some recent experiments something of this process has been captured. For example, in the context of a finding game, an adult announced her intentions to ‘find the toma’ and then searched in a row of buckets all containing novel objects. Sometimes she found it in the first bucket searched, smiling and handing the child the object. Sometimes, however, she had to search longer, rejecting unwanted objects by scowling at them and replacing them in their buckets until she found the one she wanted (again indicated by a smile and the termination of search). Children learned the new word for the object the adult intended to find regardless of (p.98) whether or how many objects were rejected during the search process (Tomasello and Barton 1994; see Tomasello 2000, for a review of other similar studies). Indeed, a strong argument can be made that children can only understand a symbolic convention in the first place if they understand their communicative partner as an intentional agent with whom one may share attention—since a linguistic symbol is nothing other than a marker for an intersubjectively shared understanding of a situation (Tomasello 1998; 2000).

Let us now turn to our cousins, the non-human primates, and to their attempts (if that is what they are) to share attention with others. Many scientists outside the field take as the
paradigm case of non-human primate communication the alarm calls of vervet monkeys. The basic facts are set out in Cheney and Seyfarth (1990). In their natural habitats in East Africa, vervet monkeys use three different types of alarm call to indicate the presence of three different types of predator: leopards, eagles, and snakes. A loud, barking call is given to leopards and other cat species, a short cough-like call is given to two species of eagle, and a ‘chutter’ call is given to a variety of dangerous snake species. Each call elicits a different escape response on the part of vervets who hear the call: to a leopard alarm they run for the trees; to an eagle alarm they look up in the air and sometimes run into the bushes; and to a snake alarm they look down at the ground, sometimes from a bipedal stance. These responses are just as distinct and frequent when researchers play back previously recorded alarm calls over a loudspeaker, indicating that the responses of the vervets are not dependent on their actually seeing the predator. On the surface, it seems as if the caller is directing the attention of others to something they do not perceive or something they do not know is present. These alarm calls would thus seem to be referential and therefore good candidates for precursors to human language.

But several additional facts argue against this interpretation. First, no ape species has such specific alarm calls (Cheney and Wrangham 1987). Since human beings are most closely related to apes, it is not possible that vervet monkey alarm calls could be the direct precursor of human language unless apes at some point used them also. Indeed, the fact is that predator-specific alarm calls are used by a number of non-primate species—from ground squirrels to domestic chickens—who must deal with multiple predators requiring different types of escape responses (Marler 1976), although no one considers these as direct precursors to human language. Second, vervet monkeys do not seem to use any of their other vocalizations referentially.

They use ‘grunts’ in various social situations (and some ape species have similar ‘close’ calls as well: Cheney and Seyfarth 1990), but these mainly serve to regulate dyadic social interactions such as grooming, play, fighting, and sex, not to draw attention to outside entities. Alarm calls thus are
not representative of other monkey calls and so they do not embody a generalized form of communication. Third, primate vocalizations are almost certainly not learned, as monkeys and apes raised outside of the normal social environments still call in much the same way as those who grow up in normal social environments (although some aspects of call comprehension and use maybe learned: Tomasello and Zuberbühler 2002). And one would think that language could only have evolved from socially learned and flexibly used communicative signals.

Our nearest primate relatives, the chimpanzees, actually communicate in more flexible and interesting ways with gestures rather than with vocalizations. Although they have a number of more or less involuntary postural and facial displays that express their mood, (e.g. piloerection indicating an aggressive mood and ‘play-face’ indicating a playful mood), they also use a number of gestures intentionally, that is, in flexible ways tailored for particular communicative circumstances. What marks these gestures as different from involuntary displays and most vocalizations is: (1) they are clearly learned, as different individuals use different sets of them; (2) they are used flexibly, both in the sense that a single gesture may be used in different contexts and in the sense that different gestures may be used in the same context; and (3) they are clearly sensitive to audience, as the signaller typically waits expectantly for a response from the recipient after the gesture has been produced (Tomasello et al. 1985).

In their natural communication with conspecifics, chimpanzees employ basically two types of intentional gesture. First, ‘attractors’ are imperative gestures aimed at getting others to look at the self. For example, when youngsters want to initiate play they often attract the attention of a partner to themselves by slapping the ground in front of, poking at, or throwing things at them (Tomasello et al. 1989). Because their function is limited to attracting attention, attractors most often attain their specific communicative goal from their combination with involuntary displays; for example, the specific desire to play is communicated by the involuntary ‘play-face’, with the attractor serving only to gain attention to it. The second type of intentional gestures are ‘incipient actions’ that have become ritualized into gestures (see
Tinbergen 1951 on ‘intention-movements’). These gestures are also imperative in function, but they communicate more directly what specifically is desired. For example, play hitting is an important part of the rough-and-tumble play of chimpanzees, and so many individuals come to use a stylized ‘arm-raise’ to indicate that they are about to hit the other and thus initiate play. Many youngsters also ritualize signals for asking their mother to lower her back so they can climb on, for example, a brief touch on the top of the rear end, ritualized from occasions on which they attempt to push her rear end down. Interestingly, in using their gestures chimpanzees demonstrate an understanding that the bodily and perceptual orientation of the recipient is an important precondition for the gesture to achieve its desired goal; for example, they use their visually based gestures only when the recipient is already looking at them (Tomasello et al. 1994).

Nevertheless, chimpanzees still do not use their gestures referentially. This is clear because (1) they almost invariably use them in dyadic contexts—either to attract the attention of others to the self or to request some behaviour of another toward the self (e.g. play, grooming, sex)—not triadically, to attract the attention of others to some outside entity; and (2) they use them exclusively for imperative purposes to request actions from others, not for declarative purposes to direct the attention of others to something simply for the sake of sharing interest in it or commenting on it. Thus, perhaps surprisingly, chimpanzees do not point to outside objects or events for others, they do not hold up objects to show them to others, and they do not even hold out objects to offer them to others (Tomasello and Call 1997).

A number of scholars have recently cautioned against using human language as an interpretive framework for non-human primate communication (Owings and Morton 1998; Owren and Rendall 2001). According to these theorists, non-human primate communicative signals are not used to convey meaning or to convey information or to refer to things or to direct the attention of others, but rather to affect the behaviour of others directly. If this interpretation is correct—and it is certainly consistent with the facts outlined above-then...
the evolutionary foundations of human language lie in the attempts of individuals to influence the behaviour of conspecifics, not their mental states. Attempting to influence the attention and mental states of others is a uniquely human activity, and so must have arisen only after humans and chimpanzees split from one another some six million years ago.

To summarize, what we are seeing at 9-12 months of age in human infants is the ontogenetic emergence of the species-unique social-cognitive adaptation that made possible human culture and, as a special case of that, human symbolic communication. Other non-human primates do not seem to have this same adaptation. Not only do non-human primates not seem to use linguistic or other types of symbols, they do not even use non-linguistic means of directing and sharing attention with others. Although it is unknown when the social-cognitive adaptation that enabled symbolic communication emerged in human evolution, one plausible hypothesis is that it emerged only very recently with modern humans, and that it was indeed this very adaptation that enabled them to out-compete other hominids (Tomasello 1999).

Grammar

It is currently popular to believe that grammar also (or perhaps only grammar) derives from a species-unique biological adaptation. Discussion of this difficult issue would take us far afield into theoretical linguistics, but some key points can nevertheless be made. The proposal that there is a biologically determined Universal Grammar—which contains specific linguistic content—rests crucially on the hypothesis that there are indeed contentful similarities, or identities, among the grammatical structures of all of the languages of the world. As we investigate more and more of the world’s 6000+ languages, this hypothesis is proving more and more difficult to maintain. Of course we can take the grammar of Standard Average European and impose it on these other languages. But when we look at, for example, Austronesian languages, on their own terms, we find that they work in some quite unexpected ways: they simply do not have some categories and constructions that appear in European...
languages, and they of course have some of their own categories and constructions as well (see Dryer 1997; Croft 2002). I do not mean to imply that there are no linguistic universals; of course there are. But these do not consist of specific linguistic categories or constructions; they consist of general communicative functions such as reference and predication, or cognitive abilities such as the tendency to conceptualize objects and events categorically, or information-processing skills such as those involved in dealing with rapid vocal sequences (Tomasello 1995).

The question thus arises: if grammatical structures do not come from the genes, where do they come from? The answer is quite well known to typolo-gists and historical linguists: from processes of grammaticalization operating (p.102) over historical time. These processes take loose discourse sequences, comprising linguistic symbols for concrete items of experience such as objects and actions, and turn them into coherent grammatical constructions with various specialized symbols that perform grammatical functions with respect to these concrete symbols, such as marking case, tense, or constituency. And this is a point that cannot be stressed too much (although I will not have time to elaborate it here): grammatical constructions are themselves symbolic. To take a very mundane example, if I say *X floosed Y the Z*, native speakers of English will immediately understand some sort of event or transfer—even without the aid of any recognizable content words. As Adele Goldberg has demonstrated in her 1995 book, and as Ron Langacker (1987; 1991) has been arguing for years, grammatical constructions are kind of linguistic gestalts that themselves function symbolically. Children hear and learn them in the same basic way (with some twists of course) that they hear and learn lexical items. Children’s contribution is that they then discern patterns across the various grammaticalized utterances they hear, and thus form the kinds of linguistic abstraction that become the hallmark of mature linguistic competence. Let us look briefly at each of these two time-frames—historical and ontogenetic—in just a bit more detail.
Each of the thousands of languages of the world has its own inventory of linguistic symbols and constructions that allow its users to share experience with one another symbolically. This inventory of symbols and constructions is grounded in universal structures of human cognition, human communication, and the mechanics of the vocal-auditory apparatus—and so all languages share some features—but the particularities of specific languages each have their own cultural histories. These particularities come from differences among the various peoples of the world in the kinds of things they think it important to talk about and the ways they think it useful to talk about them—along with various historical ‘accidents’, of course. The crucial point for current purposes is that all the symbols and constructions of a given language are not invented at once, and once invented they often do not stay the same for very long. Rather, linguistic symbols and constructions evolve and change and accumulate modifications over historical time, as human beings use them with one another and adapt them to changing circumstances.

The most important dimension of the historical process in the current context is grammaticalization or syntacticization, which involves loose and redundantly organized discourse structures congealing into tight and less redundantly organized syntactic constructions (see Traugott and Heine 1991a; 1991b; Hopper and Traugott 1993, for some recent research). For example, (1) loose discourse sequences such as He pulled the door and it opened may become syntacticized into He pulled the door open (a resultative construction); (2) loose discourse sequences such as My boyfriend… He plays piano… He plays in a band. may become My boyfriend plays piano in a band; (3) a sequence such as My boyfriend… He rides horses… He bets on them. may become My boyfriend, who rides horses, bets on them; and (4) complex sentences may also derive from discourse sequences of initially separate utterances, as in I want it… I buy it. evolving into I want to buy it. In the process, free-standing, contentful words often turn into grammatical morphemes (e.g. auxiliaries, prepositions, tense markers, case markers), as a kind of ‘glue’ that holds the new construction together. Children now learn these
constructions as symbolic wholes, which also have functionally significant constituent parts.

Systematic investigation into processes of grammaticalization and syntacticization is still in its infancy; indeed, the suggestion that languages may have evolved from structurally simpler to structurally more complex forms by means of processes of grammaticalization and syntacticization is somewhat speculative—these processes are most often thought of by linguists as sources of change only. But grammaticalization and syntacticization are able to effect serious changes of linguistic structure in relatively short periods of time—for example, the main diversification of the Romance languages took place during some hundreds of years—and thus there is no reason why they could not also work to make a simpler language more complex syntactically in some thousands of years. Exactly how grammaticalization and syntacticization happen in the concrete interactions of individual human beings and groups of human beings, and how these processes might relate to the other processes of sociogenesis by means of which human social interaction ratchets up the complexity of cultural artefacts, requires more psychologically based linguistic research into processes of linguistic communication and language change.

Turning now to ontogeny, the essential point—made by Chomsky many years ago—is that children do not hear linguistic abstractions, but only concrete utterances; they must supply the abstractions themselves. But how they do this is a point of dispute. The best-known answer—first proposed by Chomsky and more recently popularized by Pinker (1994) and others—is that children do not have to learn or construct abstract syntactic structures, but rather already possess them as a part of their innate language faculty (Universal Grammar). Recent research suggests, however, that most of young children’s early language is not based on adult-like abstractions, innate or otherwise. For example, in a detailed diary study Tomasello (1992) found that most of his English-speaking daughter’s early multi-word speech revolved around specific verbs and other predicative terms. That is to say, at any given developmental period each verb was used in its own unique set of utterance-level schemas, and across
developmental time each verb began to be used in new utterance-level schemas (and with different Tense Aspect Modality morphology) on its own developmental timetable, irrespective of what other verbs were doing during that same time period. There was thus no evidence that once the child had mastered the use of, for example, a locative construction with one verb she could then automatically use that same locative construction with other semantically appropriate verbs. Generalizing this pattern, Tomasello (1992) hypothesized that children’s early grammars could be characterized as an inventory of verb-island constructions (utterance schemas revolving around verbs), which then defined the first syntactic categories as lexically based things such as ‘hitter’, ‘thing hit’, and ‘thing hit with’ (as opposed to subject/agent, object/patient, and instrument; see also Tomasello and Brooks 1999).

Lieven et al. (1997; see also Pine et al. 1998) found some very similar results in a sample of twelve English-speaking children, and a number of systematic studies of children learning languages other than English have also found basically item-based organization. For example, in a study of young Italian-speaking children Pizzuto and Caselli (1992; 1994) found that of the six possible person–number forms for each verb in the present tense, about half of all verbs were used in one form only, and an additional 40 per cent were used with two or three forms. Of the 10 per cent of verbs that appeared in four or more forms, approximately half were highly frequent, highly irregular forms that could only have been learned by rote, not by application of an abstract rule. In a similar study of one child learning to speak Brazilian Portuguese, Rubino and Pine (1998) found adult-like subject-verb agreement patterns only for the parts of the verb paradigm that appeared with high frequency in adult language (e.g. first person singular), not for low-frequency parts of the paradigm (e.g. third person plural). The clear implication of these findings is that Romance-speaking children do not master the whole verb paradigm for all their verbs at once, but only master some endings with some verbs—and often different ones with different verbs. It should also be noted that syntactic overgeneralization errors such as Do not
fall me down—which might be seen as evidence of more
general and categorical syntactic knowledge—are almost
never produced before about 3 years of age (see Pinker 1989).

Finally, experiments using novel verbs have also found that
young children’s early productivity with syntactic
constructions is highly limited. For example, Tomasello and
Brooks (1998) exposed 2–3-year-old children to a novel verb
used to refer to a highly transitive and novel action in which
an agent was doing something to a patient. In the key
condition the novel verb was used in an intransitive sentence
frame such as The sock is tamming (to refer to a situation in
which, for example, a bear was doing something that caused a
sock to ‘tam’—similar to the verb roll or spin). Then, with novel
characters performing the target action, the adult asked
children the question: What is the doggie doing? (when the
dog was causing some new character to tam). Agent questions
of this type encourage a transitive reply such as He’s tamming
the car—which would be creative since the child has heard
this verb only in an intransitive sentence frame. The outcome
was that very few children produced a transitive utterance
with the novel verb, and in another study they were quite poor
at two tests of comprehension as well (Akhtar and Tomasello
1997). It is important that 4–5-year-old children are quite
good at using novel verbs in transitive utterances creatively,
demonstrating that once they have acquired more abstract
linguistic skills children are perfectly competent in these tasks
(Pinker et al. 1987; Maratsos et al. 1987; see Tomasello 2000b
for a review). Finally, Akhtar (1999) found that if 2.5–3.5-year-
old children heard such things as The bird the bus meeked,
when given new toys they quite often repeated the pattern and
said such things as The bear the cow meeked—only
consistently correcting to canonical English word order at 4.5
years of age. This behaviour is consistent with the view that
when 2–3-year olds are learning about meeking they are just
learning about meeking; they do not assimilate this newly
learned verb to some more abstract, verb-general linguistic
category or construction that would license a canonical
English transitive utterance.
This general approach may be extended to more complex structures. For example, Diessel and Tomasello (2001) looked at seven children’s earliest utterances with sentential complements and found that virtually all of them were composed of a simple sentence schema that the child had already mastered combined with one of a delimited set of matrix verbs (see also Bloom 1992). These matrix verbs were of two types. First were epistemic verbs such as think and know. In almost all cases children used I think to indicate their own uncertainty about something, and they basically never used the verb think in anything but this first-person, present tense form; that is, there were virtually no examples of He thinks ..., She thinks ..., etc., virtually no examples of I do not think..., I can’t think..., etc., and virtually no examples of I thought..., I didn’t think..., etc. And there were almost no uses with a complementizer (virtually no examples of I think that...). It thus appears that for many young children I think is a relatively fixed phrase meaning something like Maybe. The child then pieces together this fixed phrase with a full sentence as a sort of evidential marker, but not as a ‘sentence embedding’ as it is typically portrayed in more formal analyses. The second kind of matrix verbs are attention-getting verbs like Look and See in conjunction with full finite clauses. In this case, children use these ‘matrix’ verbs almost exclusively in imperative form (again almost no negations, no non-present tenses, no complementizers), suggesting again an item-based approach not involving syntactic embedding.

A similar story may be told with respect to children’s utterances with relative clauses (Diessel and Tomasello 2000). Virtually all of English-speaking children’s earliest relative clauses, in the period before 3-3.5 years of age, occur in presentational constructions of the type: Here’s the chair that broke, There’s the drink I want, and It’s the toy that spins. Each consists of (1) a main clause with the verb to be that is a highly practised sequence for children of this age (i.e. children have said by this time many thousands of times Here’s theX, There’s theX, and It’s theX: Lieven et al. 1997), and (2) a relative clause, usually with an intransitive verb, that conveys new information about the topic introduced with the
presentational main clause. Relative constructions of this type thus express a single proposition, with topic introduced in a common topic-introducing construction and comment expressed in the newly learned relative clause construction. After 3 or 3.5, the children began to use more complex relative constructions in which a relative clause, including an intransitive or transitive verb, was attached to a noun in a fully-fledged main clause. It is only at this point that we may accurately speak of children producing utterances with subordinate relative clauses.

The point is that young children are learning the specific linguistic items and constructions of the language they hear around them. Initially, they do not operate on the basis of any linguistic abstractions, innate or otherwise. Fairly quickly, however, they find some patterns in the way concrete nouns are used and form something like a category of noun, but schematization (p.107) across larger constructions goes more slowly. The process of how children find patterns in the ambient language and then construct categories and schemas from them is not well understood at this point. But some progress has recently been made.

Children begin the abstraction process first by creating ‘slots’ in otherwise item-based schemas (Tomasello et al. 1997). It is not known precisely how they create these slots, but one possibility is that they observe adult speech variation in that utterance position and so induce the slot on the basis of ‘type frequency’ (Langacker 1988; Krug 1998; Bybee and Schiebman 1999). The nature and extent of type variation needed for different kinds of productivity is not known at this time, and indeed after a certain point in development it may be that type variation in the slots of constructions becomes less important as these slots come to be more precisely defined functionally. Another possibility—not mutually exclusive but rather complementary to the above—is that abstract constructions are created by a relational mapping across different verb island constructions (Gentner and Markman 1997). For example, in English the several verb island constructions that children have with the verbs give, tell, show, send, and so forth, all share a ‘transfer’ meaning
and they all appear in a structure: NP+V+NP+NP (identified by the appropriate morphology on NPs and VPs). Children may thus make constructional analogies based on similarities of both form and function: two utterances or constructions are analogous if a ‘good’ structure mapping is found both on the level of linguistic form and on the level of communicative function. Precisely how this might be done is not known at this time, but there are some proposals that a key element in the process might be some kind of ‘critical mass’ of exemplars, to give children sufficient raw material from which to construct their abstractions (Marchman and Bates 1994).

One relevant study of the early stages is that of Childers and Tomasello (2001). They trained 2 2.5-year-old English-speaking children with English transitive verbs of varying degrees of familiarity to children of this age. More importantly, they also varied the type variation in the nominal slots around these verbs in the utterances the children heard. Thus, some children heard these verbs with full nouns only in the slots; for instance, as new characters were used to act out the event they heard *The bunny's striking the tree* (repeat), *The bear's striking the cat* (repeat), and so forth—and they heard a similar pattern for sixteen different verbs. Other children heard these same verbs with both pronouns and full nouns in the slots. That is, for each pair of characters they heard something like *The bear's striking the tree. He's striking it*, with the pronouns used across all models (and across all sixteen verbs) being *He's VERBing it*. These latter models thus provided children both with a familiar utterance frame (*He's VERBing it*) and with type variation, as different nouns were used as well. And indeed it was models of this latter type that best facilitated children’s later ability to use a nonce verb (e.g. *dacking*) in a transitive frame (which they mostly did with pronouns).

Overall, the main point is that young children begin by imitatively learning specific pieces of language in order to express their communicative intentions. As they attempt to comprehend and reproduce the utterances produced by mature speakers—along with the internal constituents of those utterances—they come to discern certain patterns of language
use, and these patterns lead them to construct a number of different kinds of (at first very local) linguistic categories and schemas. As with all kinds of categories and schemas in cognitive development, the conceptual ‘glue’ that holds them together is function: children categorize together things that do the same thing communicatively (see Tomasello 1992 on functionally based distributional analysis). For people who are generally sceptical that structures as complex and abstract as those embodied in the syntax of natural languages could evolve historically, we need only to point to other cultural products, such as algebra. Although, of course, the analogy with language is not perfect, algebra is a cultural product that is clearly not in the genes (the majority of cultures and people in the world do not know or use algebra and indeed we know its history). Of course human beings are quantitatively inclined biologically, but the specific structures of algebra are a product of a particular historical evolution. Languages are the same way.

Conclusion

I am afraid that I have no real evolutionary fairy tale with which to conclude. At some point, hominids were communicating with one another in the typical great ape fashion—something like modern-day chimpanzee gestures and vocalizations, perhaps. For some reason—I know not what—the individuals of one population began to understand one another as intentional agents whose attention and other psychological states to the outside world could be actively followed into, manipulated, and shared. This enabled the conventionalization—symbolization, if you prefer—of a set of communicative behaviours premised on this intersubjective understanding. Of course a good candidate for this special population is modern humans—but I really have no idea.

One obvious implication of this scenario is that there was no specific adaptation ‘for’ symbolic or linguistic communication. The adaptation was ‘for’ a particular kind of social cognition—understanding others on analogy with the self—and symbols then developed as a kind of natural consequence. When you know that someone else has psychological experiences like
your own, it is just natural for an intelligent primate to want to manipulate those states for various cooperative and competitive purposes.

Grammatical constructions then emerged from discourse patterns over historical time with no further biological adaptations—except perhaps further adaptations of the vocal-auditory apparatus to enable the ever-faster processing of real-time speech, which may then have had grammatical consequences. When children grow up in the midst of utterances embodying these constructions, they find their own patterns and make their own abstractions—which underlie the awesome creativity of mature linguistic competence and also some of the changes in languages that occur across generations historically.

And so we may see the origins and emergence of human language as one part of the much larger process of human culture, that is, as one more instance of what I have in other contexts called ‘the ratchet effect’. Like many cultural products language is a complex outcome of human cognitive and social processes taking place in evolutionary, historical, and ontogenetic time. And different aspects of language—for example, symbols and grammar—may have involved different processes at different evolutionary times.

**FURTHER READING**

There is a large and complex philosophical literature on the nature of symbols. In the modern context this begins with the work of the philosopher Charles Peirce, whose work is not so very accessible. He is the first to distinguish, for example, between icons, indices, and symbols. One reasonable collection of some of his papers is Buchler (1955). An interesting extension of his ideas may be found in Suzanne Langer’s (1957) *Philosophy in a New Key*. The work of Ernst Cassirer (1944) on the *Philosophy of Symbolic Forms* is also useful, and Nelson Goodman’s (1978) work has also been influential. In an evolutionary context, the main theoretical work is that of Sebeok (1990). In a developmental context, the work of Piaget (1962) and Werner and Kaplan (1963) are classic. Sinha (1988) also is very useful.
The main sources for learning more about processes of grammaticalization are Traugott and Heine (1991a; 1991b), Hopper and Traugott (1993), Heine (1991), Keller (1994), Trask (1996), and Croft (2000). Grammaticalization theory is well established for explaining language change, but its application to language origins has not been systematically pursued—for the obvious reason of a lack of data about exactly what kinds of thing people said early in the process.