5

The usage-based theory of language acquisition

Michael Tomasello

5.1 Introduction

The usage-based approach to linguistic communication may be summarized in the two aphorisms:

- meaning is use
- structure emerges from use

'Meaning is use' represents an approach to the functional or semantic dimension of linguistic communication. It originated with Wittgenstein (1953) and other pragmatically based philosophers of language, who wanted to combat the idea that meanings are things and instead focus on how people use linguistic conventions to achieve social ends. 'Structure emerges from use' represents an approach to the structural or grammatical dimension of linguistic communication. It is implicit in the work on grammaticalization and language change of many historical linguists, and has been made explicit by Langacker (1987, 2000) and other usage-based linguists, who want to combat the idea of a wholly formal grammar devoid of meaning and instead focus on how meaning-based grammatical constructions emerge from individual acts of language use.

Drawing on the work of many other researchers, Tomasello (2003) proposes a usage-based theory of language acquisition. Paralleling the two aphorisms above, the proposal is that children come to the process of language acquisition, at around one year of age, equipped with two sets of cognitive skills, both evolved for other, more general functions before linguistic communication emerged in the human species:

- intention-reading (functional dimension)
- pattern-finding (grammatical dimension)

'Intention-reading' is what children must do to discern the goals or intentions of mature speakers when they use linguistic conventions to achieve

social ends, and thereby to learn these conventions from them culturally. Intention-reading – including skills of joint attention – is the central cognitive construct in the so-called social-pragmatic approach to language acquisition (which is most often used in the study of word learning; Bruner 1983, Nelson 1996, Tomasello 1992, 2000d, 2001). 'Pattern-finding' is what children must do to go productively beyond the individual utterances they hear people using around them to create abstract linguistic schemas or constructions. As a summary term for such things as categorization, analogy and distributional analysis, pattern-finding is the central cognitive construct in the so-called usage-based approach to the acquisition of grammar (Goldberg 1995, 2006, Tomasello 2000a, 2003).

These theoretical positions on the functional and grammatical dimensions of language use and acquisition are minority positions in the field. Essentially, they represent the view that the pragmatics of human communication is primary, both phylogenetically and ontogenetically, and that the nature of conventional languages – and how they are acquired – can only be understood by starting from processes of communication more broadly. In this chapter I provide a synoptic account of the usage-based approach to language acquisition, in both its functional and grammatical dimensions.

5.2 Prelinguistic communication

In the usage-based view one must always begin with communicative function, and it turns out that human infants communicate in some fairly sophisticated ways before they have acquired any linguistic conventions (see Goldin-Meadow Ch. 9). For example, almost all infants communicate by pointing before they have acquired any productive language, and many also use some kind of iconic or conventionalized gestures as well. Interestingly and importantly, other animal species, including our nearest primate relatives, do not communicate with conspecifics in these ways. This suggests that human pointing and other gestures may already embody forms of social cognition and communicative motivation that are unique to the species, and that are necessary as a first step on the way to linguistic conventions both phylogentically and ontogenetically (Tomasello in press).

The interesting thing about pointing is that there is almost no information *in* the gesture itself; it basically says 'look in that direction and you'll know what I mean'. So where does the meaning come from? One can say it comes from context, but this has a very special significance with respect to human communication; specifically, it means mutually understood context. One person could point for another in exactly the same way to exactly the same clock on the wall, for example, and mean everything from 'what a beautiful clock' to 'our friend is late', depending only on their shared

experience and attention prior to the pointing act. From their earliest communicative pointing, infants understand and produce pointing gestures in the context of some such joint attentional frames or common conceptual ground (Tomasello *et al.* 2007). For example, if an 18-month-old girl is engaged in cleaning up toys with an adult, and the adult points to a toy across the room, she will fetch it and clean it up also – assuming that the adult pointing gesture is relevant to their shared activity. But if another person enters the room and points to the exact same toy in the exact same way at a comparable moment, even though the infant herself has been engaged in cleaning up (with the first adult), she does not interpret this pointing gesture as relevant to her own activity egocentrically and so she does not clean up the toy but instead shares attention to it declaratively or gives it to the new adult (Moll *et al.* in press).

Even young infants do not just communicate about what they understand of the world, but about the shared understandings they have with other potential communicative partners. Infants have the ability to construct such shared understandings – in the form of specific formats, scripts, routines or joint attentional frames in specific interactive contexts – from around the first birthday, and these structure their earliest intentional communication (Bruner 1983, Tomasello 1988). The cognitive aspect of these joint attentional frames comprises precisely those conceptualizations that will later structure young children's complex utterances: agents acting on patients, agents giving things to others, objects being in locations or moving to locations, objects changing states, people in various psychological states and so forth. Importantly, when children communicate in specific instances of such situations or events, they comprehend both their role and the role of the communicative partner. For example, in the diary observations of Carpenter *et al.* (unpublished data) a 14-month-old boy on two different occasions wants his chair pushed up to the dining room table in preparation for mealtime. On one occasion he and his mum are standing next to the table and so he points to the chair; on another occasion he and his mum are standing next to the chair and so he points to the table. This suggests that this child already has some understanding – which he knows he shares with his mum – about preparations for mealtime, where his chair goes at the table, and so forth, that serve as a kind of background topic for the communicative act. He then highlights for his mum, by pointing, the aspect of the situation he wants her to focus on – the one that is new for her – so that she can discern his communicative intention (that the chair be placed under the table in its usual place). On other occasions, with a different joint attentional frame as common ground, it is easy to imagine that this child might point to his chair wanting to be placed in it, or point to the empty space at the table simply to indicate dispassionately that the chair that is normally there is missing (and indeed the Carpenter *et al.* observations include several from prelinguistic children indicating absent referents; see also Lizskowski *et al.* 2007).

In terms of communicative motives, it is well established that infants point for both imperative and declarative motives before language (Bates *et al.* 1979, Carpenter *et al.* 1998). Recent research has even documented that 12-month-old infants point helpfully to inform others of things they are ignorant about. For example, if the mother is searching for something and the child knows where it is, even 12-month-old infants will inform her of this with no desire for the object themselves (Lizskowski *et al.* 2006). The imperative, declarative and informative motives underlying infants' prelinguistic communication are of course exactly the same motives that will structure their early language in the coming months.

Infants' prelinguistic gestural communication, therefore, already includes a species-unique ability to construct with others various kinds of joint attentional common ground to serve as background topic for the attention-directing act of pointing – comprising such things as agents, locations, objects, etc. – as well as species-unique motives for communicating (declarative and informative) that are the exact same motives with which they will use their earliest language. Indeed, many of young children's earliest uses of language are actually accompanied by pointing or other gestures, and these partition the communicative intention in ways that demonstrate the equivalence of gesture and language from a communicative point of view; for example, the child might point to the door while saying "Daddy" to indicate what he might later indicate with "Daddy leave" or some such (Iverson & Goldin-Meadow 2005). In general, prelinguistic communication paves the way for the acquisition of the 'arbitrary' linguistic conventions that infants use, initially, in exactly the same kinds of situations, for exactly the same kinds of communicative motives, as their early gestures.

5.3 Utterances and words

When we turn to children's early linguistic communication, the most basic unit of linguistic experience, and the one with which children begin, is not the word but the utterance. An utterance is the smallest unit in which a person expresses a complete communicative intention – that is, an intention that another person attend to something within the joint attentional frame and so do something as a result – and it thus corresponds to prelinguistic communicative acts such as pointing. Like an act of pointing, an utterance is used to both direct a recipient's attention to something referentially, and also to express a communicative motive (imperative, declarative, informative and others), typically through some form of emotional expression in the face and/or voice. When the child either comprehends or produces an utterance such as 'Birdie!' (to point it out) or 'Hold!' (to request), he or she understands a full communicative act, comprising both reference and motive – even though the form

is simply a single adult word expressed with a certain emotion. These so-called holophrases are thus already, in a very simple way, composite structures.

When an adult speaks to him or her, then, what the child is attempting to do most urgently is to comprehend the overall communicative intention behind the utterance; what does the adult intend for me to attend to and to do in the joint attentional situation? At the same time, he or she is also attempting to determine the communicative function of particular constituents within the utterance. This is a kind of 'blame assignment' procedure in which the child attempts to determine the functional role of a constituent in the utterance as a whole. This requires that the child determine, to some degree of specificity, the communicative intention of the whole utterance; one cannot determine a novel sub-function without knowing something about the overall function. Presumably, particular utterance constituents such as words are most easily identified – and emerge as independent units – when the same phonological form appears in different utterances over time with some functional consistency. Thus, if the child hears 'There's the ball', 'Gimme my ball', 'The ball's rolling', 'The ball's bouncing', 'I want a ball', 'Throw the ball', 'That ball's Jeffery's', 'Where's your ball?', etc., the word *ball* comes to exist as a potential utterance constituent for future use when the child needs to indicate one of a certain class of objects as one sub-function of an utterance. One thing that facilitates this process is if the adult stresses the key word, as an indication of its referential newness, and its associated referent is indeed new to the situation (Grassman & Tomasello 2007).

As a non-linguistic example, a young girl may see her father use a stapler and understand that his goal is to staple together two pieces of paper. In some cases, the girl may understand also that the sub-goal/function of placing the papers inside the stapler's jaws is to align them with the stapling mechanism inside the stapler, and that the sub-function of pressing down on the stapler is to eject the staple through the two papers – with both of these sub-functions being in the service of the overall goal of attaching the two sheets of paper. The girl does not need to understand all of this to mimic an adult stapling papers with the same stapler over and over again (analogy: child can say "There-ya-go" over and over again without understanding its internal constituents). But to the extent that the girl does not understand these sub-functions, she will be lost when she encounters some new stapler in which the sub-functions are effected by a different means, for example, one whose stapling mechanism does not require pressing down but rather squeezing. Only to the extent that the girl understands the relevant sub-functions, will she be able to adapt to new situations creatively by, for example, adjusting her behaviour to effect the same outcome with the new stapling mechanism. In the same way, the child may hear an adult say "I stapled your papers" and comprehend not only the utterance and its overall communicative intention, but also, for

example, the words *I* and *stapled* and their communicative sub-functions in the utterance (the contributions they are making to the utterance as a whole), along with the phrase *your papers* and its communicative sub-function in the utterance (and the sub-sub-functions of *your* and *papers*). As in the case of the stapler, it is only if the child performs some kind of blame assignment that she will be able to comprehend the constituent linguistic elements in a deep enough way to enable her in the future to use them creatively in novel utterances (Tomasello 2003).

This is the way children learn words. That is, children do not try to learn words directly; they try to comprehend utterances and in doing so they often must comprehend a word in the sense of determining the functional role it is playing in the utterance – and they see commonalities in this functional role across utterances. The lexicon, as it were, is thus only an emergent phenomenon in the sense of Bybee (1998). This is true despite the fact that the process is sometimes obscured in Western middle-class culture because parents and children often establish highly frequent utterance schemas for naming objects (e.g. 'That's a __'. 'It's a __', 'Here's the __', etc.). Children understand quite well the overall function of these utterances as well as the function of the open slot, with the new word in the slot always serving to name the new object in the situation. This gives the impression that what children are doing is mapping a single word onto a single object or action, or concept thereof, as in most theories of word learning (e.g. Bloom 2000, Markman 1989). But if 'mapping' means simply associative learning, this is clearly not how things work. Children are attempting to understand how the adult is using an utterance (and its constituents as sub-elements) to direct their attention. The process is not one of association or mapping but of intention-reading and blame assignment.

We may use children's learning of new word in an experiment as an example. Akhtar and Tomasello (1996) had an adult set up a joint attentional game with 24-month-old children in which a novel action was performed always and only with a particular toy character on a particular substrate (e.g. Big Bird on a swing, with other character–action pairings demonstrated as well). She then picked up Big Bird and announced "Let's meek Big Bird", but the swing was nowhere to be found – so the action was not performed. Children thus never saw the new word *meek* paired with the corresponding action. But later, when the adult handed them a new toy and told them to 'Meek it', they searched for (and found) the swing and used it to swing the new character, thus demonstrating their understanding of the action intended. The only way they could do this was to understand the adult's intentions with respect to the key objects and actions in this jointly understood situation when she originally said "Let's *meek* Big Bird." – and something of the particular intentions behind the use of *meek* – even though she never actually did it. That is to say, the child had to identify the aspect of the adult's overall communicative intention not

covered by the known parts of the utterance *let's* and *Big Bird* and connect it to the unknown word *meek*. To learn a new word, children must extract it from a larger utterance and connect it with the relevant aspect of the joint attentional frame they share with the adult.

In many ways this process is even clearer for word types other than nouns and verbs for concrete objects and actions. Thus, many function words can **only** be learned through efforts to isolate their functional contribution in some larger and less predictable set of phrases. For example, Tomasello (1987) reports that his daughter learned the preposition *of* from such expressions as *piece of ice, piece of bread, scared of that*, and *scared of monsters*. It is hard to conceive of any method of acquisition here other than some process of extracting *of* from larger expressions and attempting to discern its function in the overall utterance. Levy and Nelson (1994) make a similar argument about children's earliest uses of causal and temporal terms as *because, so, since, and, but, before* and *if*. And, of course, there can be no question of mapping or association when what is involved is not learning a word per se, but rather learning which referential term of several to choose for a given referent – for example, *the chair* or *that chair in my room* or *it* – in different communicative situations. Learning to make these pragmatic choices in the conventional way – so-called referential choice – requires children to understand why a person chose one means of expression rather than another, that is, her intentions in making the choice (Matthews *et al.* 2006).

5.4 Schemas and constructions

This communication-based, usage-based way of looking at things means we cannot explain children's acquisition of grammatical competence by starting with individual words, learned in isolation, and then gluing them together with abstract meaningless rules, as in the very common 'words and rules' approach (Pinker 1999). Instead, we must begin with children's comprehension and production of whole, meaningful utterances. We then investigate how children *extract* words (with their functions) from utterances and, at the same time, how they find analogical patterns across utterances (based mainly on communicative function) and thereby *abstract* meaningful grammatical constructions.

A linguistic construction is prototypically a unit of language that comprises multiple linguistic elements used together for a relatively coherent communicative function, with sub-functions being performed by the elements as well. Consequently, constructions may vary in their complexity depending on the number of elements involved and their interrelations. For example, the English regular plural construction (N+*s*) is relatively simple, whereas the passive construction (NP *was* VERB*ed by* NP) is relatively complex. Constructions also vary in their abstractness, from abstract

constructions such as the English plural and passive, to various concrete idioms such as *kick the bucket* and *hold one's breath*. Importantly, even the most abstract constructions are still symbolic, as they possess a coherent, if abstract, meaning in relative independence of the lexical items involved (Goldberg 1995). Thus, we know the general profile of the event when we hear 'The dax got mibbed by the gazzer', even though we know none of the individual content words.

Children begin, as noted above, by producing holophrases – one unit utterances with an intonational contour expressing communicative motive. Their earliest multi-unit utterances soon form schemas or constructions, but ones that are highly concrete, not abstract (i.e. based on particular words and phrases not abstract categories). From the point of view of linguistic form, the utterance-level constructions underlying children's earliest multi-word utterances come in three types: word combinations, pivot schemas, and item-based constructions.

5.4.1 Word combinations

Beginning at around 18 months of age, many children combine two words or holophrases in situations in which they both are relevant – with both words having roughly equivalent status. For example, a child has learned to name a ball and a table and then spies a ball on a table and says, "Ball table". Utterances of this type include both 'successive single-word utterances' (with a pause between them; Bloom 1973) and 'word combinations' or 'expressions' (under a single intonational contour). The defining feature of word combinations or expressions is that they partition the experiential scene into multiple symbolizable units – in a way that holophrases obviously (by definition) do not – and they are totally concrete in the sense that they are comprised only of concrete pieces of language, not categories.

5.4.2 Pivot schemas

Beginning at around this same age, however, many of children's multi-word productions show a more systematic pattern. Often there is one word or phrase that seems to structure the utterance in the sense that it determines the speech act function of the utterance as a whole (often with help from an intonational contour), with the other linguistic item(s) simply filling in variable slot(s) – the first type of linguistic abstraction. Thus, in many of these early utterances one event-word is used with a wide variety of object labels (e.g. 'More milk', 'More grapes', 'More juice') yielding a schema such as 'More __'. Following Braine (1963), we may call these pivot schemas or constructions (see also Lieven *et al.* 1997, 2003).

Not only are pivot schemas organized only locally, but even within themselves they do not have syntax; that is, 'Gone juice' does not mean something different from 'Juice gone' (and there is no other marking to