A SENSORIMOTOR CHARACTERISATION OF SYNTAX, AND ITS IMPLICATIONS FOR MODELS OF LANGUAGE EVOLUTION

ALISTAIR KNOTT

Dept of Computer Science, University of Otago, Dunedin, New Zealand
alik@cs.otago.ac.nz

In this paper I consider the possibility that language is more strongly grounded in sensorimotor cognition than is normally assumed—a scenario which would be providential for language evolution theorists. I argue that the syntactic theory most compatible with this scenario, perhaps surprisingly, is generative grammar. I suggest that there may be a way of interpreting the syntactic structures posited in one theory of generative grammar (Minimalism) as descriptions of sensorimotor processing, and discuss the implications of this for models of language evolution.

1. An optimistic idea about how to study language evolution

One way of studying language evolution is to investigate the interface between language and sensorimotor representations in modern humans. We know that there is an interface, of course, because we can talk about what we see and do in the world. But opinions vary about how much work is involved in converting sensorimotor signals into an utterance. If language is a Fodorian module, then a lot of work is involved, because there is no overlap between the sensorimotor mechanisms which create an episode representation and the syntactic mechanisms which express it as an utterance. But many cognitive scientists now argue that syntactic mechanisms supervene to some extent on sensorimotor ones (see e.g. Rizzolatti and Arbib, 1998; Barsalou, 1999; Hurford, 2003). The more overlap there is, the less specifically linguistic machinery we need to postulate, and the simpler a story we can tell about language evolution.

In this paper, I want to be optimistic, and entertain the scenario that syntactic mechanisms overlap extensively with sensorimotor ones. While this need not be the case, it would certainly be providential for language evolution theorists. For one thing, it would make it very likely that biological specialisations for language evolved as adaptations of sensorimotor mechanisms, as has already been proposed by several theorists (see e.g. Arbib, 2005; Fadiga *et al.*, 2006). Furthermore, given that our sensorimotor capabilities are relatively similar to those of other primates (see e.g. Tootell *et al.*, 1996; Iacoboni, 2006), we are probably within our rights to use the modern primate sensorimotor system as an approximation of the preadaptive platform from which language evolved. Models of the sensorimotor system would then provide a concrete starting point for relatively detailed hypotheses about how language evolved. In short, the possibility that syntactic mechanisms overlap extensively with sensorimotor mechanisms is one which language evolution theorists should think seriously about.

What would syntactic theory look like if language were indeed strongly grounded in sensorimotor mechanisms? Given that all humans have the same sensorimotor apparatus, one thing we expect is that there should be minimal differences between the grammars of different languages around the world. This seems like a setback, because languages around the world appear to differ quite a lot. In fact, the only way of maintaining our optimistic scenario is to assume that these differences are relatively superficial, and that at some deeper level of syntactic representation, the mechanisms for generating sentences are basically the same from one language to another.

Perhaps paradoxically, if we want to explore the best possible scenario a language evolution theorist could hope for, the syntactic framework we are drawn towards is Chomskyan generative grammar. This framework assumes that language is largely the product of an innate mechanism, which operates in the same way in every language. Accordingly, the syntax of a sentence is specified at an 'underlying' level, which reflects the operation of this innate mechanism and is relatively invariant across languages, and then at a more superficial level, which expresses how the underlying representation is rendered in different languages. Chomskyan generative grammar is a hunt for underlying syntactic representations (often abstract and arcane) which permit generalisations to be expressed about the syntax of different languages. If syntax supervenes heavily on sensorimotor mechanisms, such generalisations are to be expected.

Of course Chomskyan grammarians are normally also Fodorians, holding that the mechanisms responsible for language are *specific* to language. This possibility makes an account of language evolution especially difficult, as already noted, because there is a lot of language-specific machinery to evolve. On the alternative possibility that I am considering, the commonalities between languages are due to the fact that language supervenes heavily on sensorimotor mechanisms. If this is the case, then we expect to find the kind of underlying syntactic representations posited by Chomskyan linguists—but we also predict that they can be understood as descriptions of sensorimotor processing. This is quite a long shot, since the representations are developed by linguists without any reference to sensorimotor cognition at all. But the prediction can certainly be tested. We can proceed as follows. Take a simple concrete episode, which an observer can apprehend using sensorimotor mechanisms, and formulate a model of these mechanisms. Then take a sentence which reports this episode, and determine its underlying syntactic structure, within your favourite model of generative grammar. Is there any way of interpreting the syntactic structure as a description of mechanisms in the sensorimotor model? If there is, the prediction is borne out for this sentence/episode pair. If the interpretation also extends to other sentence/episode pairs, then it becomes possible to think of the language universals proposed within generative grammar as having a sensorimotor origin. A convincing sensorimotor interpretation of underlying syntactic representations would have many implications in

linguistics. Most obviously it would open up new ways of studying these representations within the domain of sensorimotor neuroscience. But it would also be good news for language evolution theorists, for the reasons already given.

In the remainder of the paper I will report on my own investigation into the relationship between sensorimotor cognition and underlying syntactic representations. The investigation is described in detail in a book I am preparing (Knott, 2010); what I give here is a summary of the main ideas. The book focusses on a single concrete episode—a man grabbing a cup. In Section 2, I give a model of the sensorimotor processes involved in experiencing this episode. In Section 3, I give a model of the underlying syntactic structure of the the associated transitive sentence *The man grabbed a cup*, expressed within the Minimalist framework of Chomsky (1995). In Section 4 I argue that there is a natural sensorimotor interpretation of this syntactic structure, and I discuss what implications this interpretation might have for an account of language evolution, if it proves to be well-founded.

2. Outline of a sensorimotor model of a reach-to-grasp action

Research in sensorimotor cognition tends to focus on processes much smaller than the perception of a complete episode. There is a great deal of work on how individual objects are attended to and categorised, on how individual actions are executed and perceived, and on how attentional and motor processes are coupled during action execution and perception. I will begin by summarising what is known about these processes, and then outline my suggestion about how they combine during the experience of a complete episode.

Perceiving an object involves attending to it and classifying it. It is known that these processes happen in different neural pathways (Milner and Goodale, 1995), and in most models an object must be attended to before it can be categorised (see classically Treisman and Gelade 1980, and more recently Reynolds and Desimone, 1999). As regards action perception: it is fairly orthodox since the discovery of mirror neurons to assume that recognising a particular reach-to-grasp action activates the same premotor representations that are involved in executing this action. As regards the coupling of attentional and motor processes: it is well established that an agent typically attends to a target object before reaching for it (see e.g. Johansson and Westling, 2001). More recently, it has been found that observers watching a reach-to-grasp action saccade to the target object well before the agent's hand reaches it (Flanagan and Johansson, 2003). This is in line with computational models of hand action classification, which assume that the observer monitors the trajectory of the agent's hand in relation to the intended target (see e.g. Oztop and Arbib, 2002). In summary, recognition and execution of a reach-to-grasp action have much in common: in each case, the experiencer first attends to and classifies the intended target object, and then evokes a premotor action representation encoding the action being performed.

The way in which an experiencer identifies the agent of a reach-to-grasp ac-

tion depends on whether he is observing the action or performing it. In the former case, the agent must be attended to as an external object. It has recently been found that observers of a reach-to-grasp action typically fixate the agent before saccading to the target (Webb *et al.*, in press). The initial saccade to the agent allows the agent to be classified or recognised, but also provides information about the agent's intentions, which is what allows the observer to make an anticipatory saccade to the target. In the latter case, where the experiencer of the action is also the agent, the mechanism which allows him to attribute the action to himself is quite different, involving recognition of a particular configuration of the motor system, in which high-level motor plans cause physical movements (see e.g. Farrer and Frith, 2002). The first operation in any action must be one which configures the motor system for action execution rather than action observation. This operation can be studied in an ERP paradigm. The earliest cortical indication of a forthcoming voluntary action is a bilateral signal called the Bereitschaftspotential (BP; Shibasaki and Hallett, 2006). Since the BP precedes lateralised activity in premotor areas reflecting a specific action being planned, it appears to encode a relatively pure 'decision to act', rather than a particular motor plan. I suggest that this operation is the means by which an experiencer establishes himself as the agent of his own actions; and therefore that the very first sensorimotor operation in both action perception and action execution is an action of 'attention to the agent'.

The picture which emerges from the above data is that experience of a reach-to-grasp action involves a canonical sequence of sensorimotor operations—and moreover that the sequence of operations is the same for action execution and action observation. The experiencer first attends to (and classifies) the agent, then attends to (and classifies) the target, and finally monitors the action to completion. In fact, during this latter process, there is evidence that the experiencer *reattends* first to the agent and then to the target. While the action is ongoing, the temporally extended pattern of signals in the sensorimotor system is characteristic of the agent as well as of the particular action being performed. When monitoring these signals, the experiencer evokes a representation of the agent as an animate entity, which is integrated with representations of the agent as a static object (Giese, 2000). When the action is complete, the agent is grasping the target object. A grasp action is a substantive motor action, but it is also an attentional action in the haptic modality, providing the agent with a new means of characterising the location and shape of the target object. Thus the target is attended to once when the action is being prepared, and again, in a different modality, when it is completed. Note that the actions of reattention to the agent and the target during action monitoring both allow the development of cross-modal object representations. The agent is attended to first as an object and then as an animate entity; the target is attended to first as an object and then as a motor state (i.e. a Gibsonian affordance).

The sequence of sensorimotor operations involved in experiencing a reach-to-grasp action according to the above account is summarised in Figure 1. Given

Initial context	Deictic operation	Reafferent sensory state	New context
$C1$	Attend to the agent	Attending to the agent	$C2$
$C2$	Attend to the target	Attending to the target	$C3$
$C3$	Activate 'grasp' action	Re-attending to the agent	$C4$
$C4$		Re-attending to the target	

Figure 1. Sequence of sensorimotor operations involved in experience of a reach-to-grasp action

that the sequencing of operations is important, it is useful to model the process of experiencing a reach-to-grasp episode as a 'deictic routine' (Ballard *et al.*, 1997). A deictic routine is a sequence of attentional and motor operations, in which each operation brings about the sensorimotor context necessary to perform the next operation. Each item in the sequence has the same basic form: a **deictic operation** causes a transition from an **initial context** to a **new context**, generating as a side-effect a **reafferent sensory state**. Note that on this model, a reach-to-grasp episode can be stored in working memory as a planned sequence of sensorimotor operations. It thus lends itself particularly well to a 'simulationist' account of meaning (see e.g. Feldman and Narayanan, 2004; Gallese, 2005): the process of evoking a semantic episode representation can be understood as the process of internally replaying a stored sensorimotor sequence from working memory.

3. Outline of a syntactic model of transitive sentences

The syntactic framework I have adopted is that of Minimalism (Chomsky, 1995). In this framework, a sentence is represented at two syntactic levels: an underlying level of 'logical form' (LF), which is relatively invariant across languages, and a surface level of 'phonetic form' (PF). The LF of our example sentence *The man grabbed a cup* is shown in Figure 2. It is formed from applications of an abstract

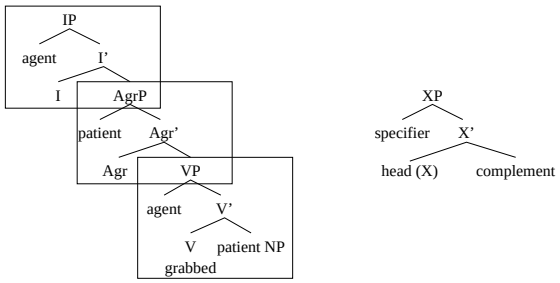

Figure 2. Left: LF structure of *The man grabbed a cup*. Right: XP schema from which it is formed.

structure called an X-bar ('XP') schema, which is shown on the right of the figure. The LF of a clause is basically a right-branching structure of XP schemas.

An XP schema contains three positions: a **head** (which in the simplest case is occupied by a word), and a **specifier** and **complement** (which are occupied by other XPs which depend on the head). For instance, the verb *grabbed* sits at the head of VP; the specifier of VP holds an NP denoting the agent of the grab action, and its complement holds an NP denoting its patient. (The internal structure of NPs is omitted in the figure.) XP schemas are also contributed by grammatical elements; in our example sentence, IP and AgrP are associated with inflections on the verb agreeing with the subject and object NP respectively.

Note that the LF structure contains two positions each for the agent and the patient. In the Minimalist account, the agent and patient are generated at the lower positions, within VP, but they must raise to the higher positions in order to be assigned a syntactic feature called 'Case'—which at a first approximation is what distinguishes between nominative and accusative pronouns. The verb must also raise, by a different mechanism called 'head movement', to the head positions associated with its inflections: first to Agr, then to I. The surface (PF) form of the sentence is 'read out' at some point during these movement operations. How this happens is relatively unconstrained; different languages have different conventions about whether the subject, object and verb are read out 'high' or 'low', which result in different basic word orders (SVO, SOV etc).

Why is the Minimalist account of a simple transitive sentence so complicated? One reason is that Minimalism has wide coverage: the mechanisms which generate *this* sentence also generate a good proportion of the other sentences in English. Any wide-coverage grammar will give a complex syntactic analysis of any given sentence. But what makes Minimalism more complex than most syntactic theories is that it attempts to define a grammar of all languages, not just of a single language. By altering the conventions about how LF structures are read out at PF, the Minimalist account should be able to model a whole space of natural languages. It is surely laudable to look for a universal mechanism underlying the languages of the world. But even so, cognitive scientists tend not to like the Minimalist notion of LF. For one thing, Minimalism has problems as a linguistic theory. Recent work in linguistics has emphasised the importance of surface patterns in language (see e.g. Goldberg, 1995; Tomasello, 2003). As Jackendoff (2002) has incisively argued, the fact that PF supervenes entirely on LF makes it hard to account for these patterns. But perhaps more importantly, Minimalists make no attempt to relate the process of constructing an LF representation to any actual cognitive process. Notoriously, it is not a 'processing model'. But then what is it??

4. A sensorimotor interpretation of LF structure

My project is to look for a way of interpreting the LF of *The man grabbed a cup* as a description of the sensorimotor processes involved in experiencing an episode in which a man grabs a cup. There is in fact an interesting isomorphism between the LF representation just described and the sensorimotor model given

in Section 2. The LF representation involves a right-branching structure of four XP schemas, associated respectively with the agent, the patient, the grab action, and again the patient. The sensorimotor model envisages a deictic routine with four phases: attention to the agent, attention to the target, activation of the 'grasp' motor programme, and reattention to the target (see Figure 1). Each XP schema has the same internal structure, and each phase in the deictic routine involves the same basic operations. In fact, I suggest we may be able to give a very general sensorimotor interpretation of an XP schema at LF, as illustrated on the left of Figure 3. Note that according to this interpretation, a right-branching sequence

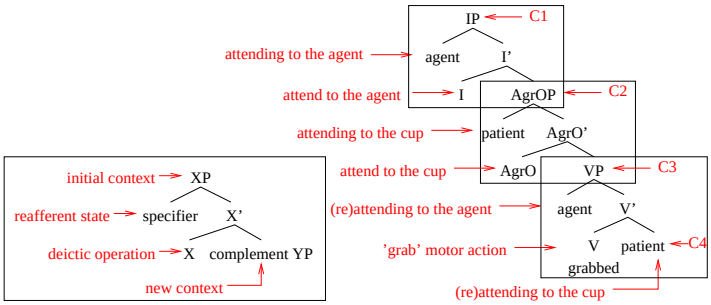

Figure 3. Left: sensorimotor interpretation of a single XP schema. Right: sensorimotor interpretation of the LF of *The man grabbed a cup*.

of XP schemas describes a sequence of deictic operations—i.e. a deictic routine. And the deictic routine which I argued for in Section 2 fits perfectly onto the LF structure of the associated sentence, as illustrated on the right of Figure 3.

This interpretation casts the Minimalist conception of LF in a completely new light. If it is legitimate, then maybe the abstract universal principles which Minimalists have derived from linguistic argumentation are not reflections of a modular language faculty, but of the fact that language is deeply grounded in sensorimotor cognition. For language evolution theorists, this is the providential scenario I began by considering. Of course, there is much work to be done to corroborate this idea. In the book I am preparing (Knott, 2010) I expand on the above analysis, arguing that there is a very natural sensorimotor interpretation of the movement of NPs from lower to higher positions in LF, and also of the raising of the verb head to higher head positions. I also argue that the sensorimotor interpretation of LF extends to several other syntactic structures: in particular, to the internal syntax of noun phrases, to the syntax of predication, and to intransitive and ditransitive sentences. I also argue that the sensorimotor interpretation of LF makes it possible to think of the Minimalist 'generative mechanism' as a model of a concrete cognitive process—one, moreover, which is involved in the generation and interpretation of

sentences. Altogether this is a very radical reinterpretation of Minimalist syntax. But my hope is that it is still recognisable by Minimalists, while at the same time having a wider relevance for researchers interested in sensorimotor processing, language processing and language evolution.

References

Arbib, M. (2005). From monkey-like action recognition to human language: an evolutionary framework for neurolinguistics. *Behavioral and Brain Sciences*, *28*(2), 105–167.

Ballard, D., Hayhoe, M., Pook, P., & Rao, R. (1997). Deictic codes for the embodiment of cognition. *Behavioral and Brain Sciences*, *20*(4), 723–767.

Barsalou, L. (1999). Perceptual symbol systems. *Behavioral and Brain Sciences*, *22*, 577–660.

Chomsky, N. (1995). *The Minimalist program.* Cambridge, MA: MIT Press.

Fadiga, L., Roy, A., Fazio, P., & Craighero, L. (2006). From hand actions to speech: evidence and speculations. In P. Haggard, Y. Rossetti, & M. Kawato (Eds.), *Sensorimotor foundations of higher cognition. attention and performance XXII* (pp. 409–434).

Farrer, C., & Frith, C. (2002). Experiencing oneself vs another person as being the cause of an action: The neural correlates of the experience of agency. *NeuroImage*, *15*, 596–603.

Feldman, J., & Narayanan, S. (2004). Embodiment in a neural theory of language. *Brain and Language*, *89*(2), 385–392.

Flanagan, J., & Johansson, R. (2003). Action plans used in action observation. *Nature*, *424*, 769–771.

Gallese, V., & Lakoff, G. (2005). The brain's concepts: The role of the sensory-motor system in conceptual knowledge. *Cognitive Neuropsychology*, *22*(3/4), 455-479.

Giese, M. (2000). Neural model for the recognition of biological motion. In G. Baratoff & H. Neumann (Eds.), *Dynamische perzeption* (pp. 105–110). Berlin: Infix Verlag.

Goldberg, A. (Ed.). (1995). *Constructions. a construction grammar approach to argument structure.* Chicago: University of Chicago Press.

Hurford, J. (2003). The neural basis of predicate-argument structure. *Behavioral and Brain Sciences*, *26*(3), 261–283.

Iacoboni, M. (2006). Visuo-motor integration and control in the human posterior parietal cortex: Evidence from TMS and fMRI. *Neuropsychologia*, *44*, 2691–2699.

Jackendoff, R. (2002). *Foundations of language: Brain, meaning, grammar, evolution.* Oxford: Oxford University Press.

Johansson, R., Westling, G., Backstrom, A., & Flanagan, J. (2001). Eye-hand coordination in object manipulation. *Journal of Neuroscience*, *21*(17), 6917–6932.

Knott, A. (2010). *Sensorimotor cognition and natural language syntax.* Manuscript, Dept of Computer Science, University of Otago. (http://www.cs.otago.ac.nz/staffpriv/alik/publications.html)

Milner, R., & Goodale, M. (1995). *The visual brain in action.* Oxford: Oxford University Press.

Oztop, E., & Arbib, M. (2002). Schema design and implementation of the grasp-related mirror neuron system. *Biological Cybernetics*, *87*, 116–140.

Reynolds, J., & Desimone, R. (1999). The role of neural mechanisms of attention in solving the binding problem. *Neuron*, *24*, 19–29.

Rizzolatti, G., & Arbib, M. (1998). Language within our grasp. *Trends in Neurosciences*, *21*, 188–194.

Shibata, H., & Hallett, M. (2006). What is the Bereitschaftspotential? *Clinical Neurophysiology*, *117*, 2341–2356.

Tomasello, M. (2003). *Constructing a language: A usage-based theory of language acquisition.* Harvard University Press.

Tootell, R., Dale, A., Sereno, M., & Malach, R. (1996). New images from human visual cortex. *Trends in Neurosciences*, *19*(11), 481–489.

Treisman, A., & Gelade, G. (1980). A feature integration theory of attention. *Cognitive Psychology*, *12*, 97–136.

Webb, A., Knott, A., & MacAskill, M. (in press). Eye movements during transitive action observation have sequential structure. *Acta Psychologica*.