Chomsky and embodied cognition: A sensorimotor interpretation of Minimalist logical form

Alistair Knott (alik@cs.otago.ac.nz)
Department of Computer Science, University of Otago, P.O. Box 56, Dunedin, New Zealand

Abstract

In this paper I argue that the Minimalist syntactic model of Chomsky (1995) may be a suitable vehicle for expressing an 'embodied' model of language, positing that language supervenes strongly on the sensorimotor system. The basic idea is that the principles which define the 'logical form' (LF) of concrete sentences, which Minimalists see as reflecting innate knowledge of language, may in fact have their origin in constraints in sensorimotor processing.
Keywords: Minimalist syntax, embodied cognition

Introduction

One of the cornerstones of Chomsykan syntax is that infants are born with innate knowledge about language—and that as a consequence, there are some syntactic principles which apply universally to all human languages. In most presentations of this idea, the innate knowledge responsible for syntactic universals is assumed to be for the most part *specialised* for language; see e.g. Chomsky (1995:167). However, the idea that syntax uses specialised neural circuitry is now somewhat open to question; Much recent evidence suggests that brain areas involved in syntactic processing also have other nonlinguistic functions, for instance in the control of actions (Novick *et al.*, 2005) or in general sequencing tasks (Dominey *et al.*, 2006). In the light of such evidence, many recent models of language have an 'embodied' flavour, proposing that language is intimately connected to the sensorimotor system (see e.g. Feldman and Narayanan, 2004; Arbib, 2005). These models are normally expressed using one of the newer 'usage-based' syntactic frameworks offered as an alternative to the Chomskyan paradigm.

My interest in the current paper is in what an 'embodied' theory of language must say about linguistic universals. It is presumably uncontroversial that all humans have the same sensorimotor system. If language supervenes on this system, as embodied theorists believe, we expect to find similarities between different languages. If language is *deeply* rooted in sensorimotor cognition, we expect *substantial* similarities between the languages of the world. To express a strong claim about the embodied nature of language, we need to adopt a syntactic model which makes correspondingly strong claims about linguistic universals.

This reasoning suggests that linguists who want to ground language in the sensorimotor system have an interest in reexamining the Chomskyan model. The syntactic universals posited in the Chomskyan account need not be reflections of a modular 'language acquisition device'. If language is closely connected to the sensorimotor system, Chomskyan universals might instead reflect properties of the sensorimotor system which we all share. Of course, exploring this idea is an inherently interdisciplinary enterprise: we need to look at the technical details of a Chomskyan account of universals, in the light of a detailed model of sensorimotor processing.

In this paper I summarise the results of a large study investigating whether a sensorimotor interpretation of Chomskyan universals can be found. The study is reported more fully in a forthcoming book (Knott, in press). The study focuses on sentences reporting one particular concrete epsiode, in which a man grabs a cup. In the first part of the study, I develop a detailed model of the sensorimotor processing involved in actually experiencing this episode, and in storing it in working memory. In the second part of the study, I introduce a syntactic model of simple sentences reporting the episode: the English sentence *The man grabs the cup*, and its equivalents in other languages. I express this model using the Minimalist syntactic framework of Chomsky (1995). In Minimalism, each sentence is represented at two levels: **PF** (**phonetic form**), which specifies how the language faculty encodes its sound, and **LF** (**logical form**), which specifies how the faculty encodes its meaning. Many syntactic universals are expressed at the level of LF: for instance, in the version of Minimalism which I adopt, sentences describing the cup-grabbing episode in different languages have a variety of different PF structures, but share the same LF structure. In the third part of the study, I argue that the LF of 'The man grabs a cup' (and equivalents in other languages) can in fact be understood as a fairly direct description of the sensorimotor processing involved in experiencing the episode it reports.

Sensorimotor processes involved in experiencing a reach-to-grasp episode

A cup-grabbing episode can be experienced in two ways: either from the perspective of the agent of the action or from that of a third-party observer. Syntax is relatively insensitive to this difference: *The man grabs a cup* can be understood as reporting the event from either perspective. If syntax supervenes on sensorimotor mechanisms, we expect the sensorimotor processing involved in actually grabbing a cup to be similar to that involved in perceiving someone else grabbing a cup. Of course, there is already some evidence to this effect: as is well known, the 'mirror system hypothesis' postulates that the neural mechanisms responsible for action recognition overlap with those responsible for action execution (see e.g. Rizzolatti, 2000). However, experiencing a cup-grabbing *episode* involves much more than just evoking an action representation. The agent and patient of the action must also be determined, and the roles they play in the action must be established. This calls on quite a wide range of sensorimotor

and cognitive mechanisms, including decision processes, attentional processes and object classification processes as well as action monitoring operations. In this section, I will outline an account of how these mechanisms interact, which emphasises their sequential organisation. My proposal is that experiencing a reach-to-grasp episode involves a series of sensorimotor operations which must occur in fairly strict sequence.

The first operation is a deployment of attention to the agent of the action. To motivate the idea that this operation has to come first, recall that our sensorimotor model must cover both action execution and action recognition. If one is executing an action, 'identifying the agent' amounts to *deciding to act*. And if one is perceiving it, then given that the mirror system must be configured differently for action execution and action recognition (see e.g. Oztop and Arbib, 2002), 'identifying the agent' has as a concomitant a decision *not to act*, but to observe instead. The operation which puts the mirror system into 'execution mode' can be thought of as an action of attention to the agent of the forthcoming action, in the sense that it establishes a representation of the agent's own body (derived from proprioception) within the system. I argue that the operation which puts the system into 'recognition mode' is also likely to be triggered by an action of attention to the agent—in this case an external agent. There must be *something* to observe, which is interesting enough to warrant the establishment of recognition mode. If the external stimulus is a reach-to-grasp action, we know empirically that observers' attention is reliably drawn first to the agent (as shown in a recent study by Webb *et al.*, 2010). In this case, information about the agent arrives visually rather than proprioceptively. In sum, whether the action is one's own or someone else's, the first operation involved in experiencing it is 'an action of attention to the agent', and a consequent sensory representation of the agent.

The second operation is an action of attention to the target of the reach-to-grasp action. If the action is one's own, it is clear one must select a target before executing a reach action towards it; there is good evidence that agents saccade to the target in the early stages of their reach movement (Johansson *et al.*, 2001). If the action is that of an observed agent, attention to the target involves inference of this agent's intentions. But again there is good evidence that observers saccade to the inferred target well before the agent's hand actually reaches it (Flanagan and Johansson, 2003), even when the target is unpredictable (Webb *et al.*, 2010). The reason for this early attention to the intended target probably relates to the way the mirror system is trained. In the standard account, an agent trains his mirror system by mapping visual representations of his own hand reaching for targets onto the motor programmes which actually control these reaches (Oztop and Arbib, 2002). Since the agent attends to the target when performing his own reach actions, he must do the same when watching those of others, so that the visual representations of observed actions are similar to those on which the mirror system was trained. In sum, whether the action is one's own or someone else's, once the agent of the action has been established, the next operation is an action of attention to the target, resulting in a sensory representation of this target.

The third operation is the monitoring and classification of the motor action taking place. Whether the action is one's own or someone else's, it is only after the target has been attended to that the observer can activate a particular motor programme representing the action. If the observer is the agent, selecting a motor programme is a matter of deciding what action to do: this can only be done after the object has been attended to, because it is only at this point that its grasp affordances are computed. If the observer is watching someone else, activating a motor programme is again a matter of inference rather than choice. The observer must monitor the trajectory of the agent's hand towards the target, and the way the hand is preshaped (Oztop and Arbib, 2002); crucially, these movements must be defined relative to the target, so they cannot be computed until the target has been attended to. In either case, action monitoring involves the activation of one particular motor programme.

The process of action monitoring has reafferent sensory consequences, just like the processes of attention to the agent and target. When we are monitoring an action, we are also unavoidably watching the agent. But the sensory representation of the agent which is activated is different from that evoked by a simple action of visual attention. When our attention is initially drawn to the agent, we represent the agent as an object with a particular shape; this is what allows us to classify the agent (as 'a man'), and in some cases also to recognise him as a specific person. When we monitor the agent's action, we represent the agent *as an agent*—in other words, as a pattern of motion. In fact, our conception of 'an agent' is a combination of static and dynamic representations: agents have characteristic shapes, but also characteristic patterns of motion. In order to form cross-modal representations of agents, it is important to attend to agents as objects while their actions are being monitored, so that the shape of an object and its pattern of motion can be bound together. There is good evidence that action recognition involves processing of form as well as of motion; see e.g. Giese (2000) for a review. Given that visual attention must be directed to the target of the action before monitoring can begin (as argued above), it appears that action monitoring involves switching some measure of attention away from the target, and back to the agent.

The final operation involved in experiencing a reach-to-grasp action involves registering that the action is complete, i.e. that the agent has successfully grasped the target. This can also be thought of in attentional terms. When we grasp an object, we bring about a change in the world (we now have the object!), but we also bring about a change in the way we *sense* the world: our sense of touch provides us with some new information about the object we are holding. In fact, grasping the cup is an action of reattention to the cup, in the haptic modality. Again, this action of reattention can be useful for the formation of a crossmodal object representation—

this time of the target object. Agents must learn a function which computes the grasp affordances of objects from their visual shapes. When we are attending to an object we are grasping, we can generate training data for this function.

In summary: synthesising evidence from a range of sources, there is some indication that experiencing a cup-grabbing episode involves a fairly well-defined sequence of sensorimotor operations. The sequence is summarised in Figure 1. In this figure, the process of experiencing a reach-to-

Initial context	Deictic operation	Reafferent sensory state	New context
$C1$	Attend to agent	Attending to agent	$C2$
$C2$	Attend to target	Attending to target	$C3$
$C3$	Activate 'grasp'	Reattending to agent	$C4$
$C4$		Reattending to target	

Figure 1: Deictic routine involved in experiencing a reach-to-grasp action

grasp episode is characterised as a **deictic routine**, drawing on Ballard *et al.*'s (1997) model of embodied cognition. The central notion in Ballard' *et al.*'s model is that of **contexts**: for an observer, a context consists of a particular internal cognitive state, paired with a particular momentary deployment of motor and attentional resources to features of his external environment. Each context enables various different **deictic operations**, which can be attentional actions, substantive motor actions, or changes of cognitive mode. Each deictic operation occurs in a particular **initial context**, and brings about a **new context**, as well as a **reafferent sensory state** carrying information about this new context. Ballard *et al.* propose that sensorimotor cognition, and perhaps cognition generally, is organised into sequences of operations of this kind. In my account, the deictic routine involved in experiencing a reach-to-grasp action has four operations, as shown in Figure 1. The routine is recursively structured: each deictic operation brings about the initial context of the next operation.

The sensoriomotor processing involved in experiencing a reach-to-grasp episode is extremely complex, both in terms of the individual mechanisms involved and their interactions. However, as Figure 1 shows, the sequential dependencies between the various operations involved can be quite compactly stated as a deictic routine, and in fact this routine holds much of the essential information about the structure of the episode. It tells us about the type of the action, and about the identity of its participants. It also tells us about the roles these participants play in the action. The first participant to be attended to is the agent: this is the participant whose actions the mirror system is configured to encode, and which is represented as a pattern of motion when the action is monitored. The second participant to be attended to is the target: this is the participant whose affordances are computed, and which ends up being established in the haptic modality. We know that expressing an observed episode as a sentence involves a huge degree of compression, converting a rich multimodal signal into a simple symbolic form. The sequential structure of the sensorimotor operations which generate the experience are a promising basis for making the necessary compression.

I argue that the deictic routine used to experience a reach-to-grasp episode is also used to store the episode in working memory. Humans and other primates are good at holding sensorimotor sequences in working memory: in macaque, there is good evidence that both attentional sequences (sequences of eye movements) and motor sequences (sequences of hand movements) are prepared in dorsolateral prefrontal cortex, and quite a lot is known about the way these sequences are represented (see e.g. Barone and Joseph, 1989; Averbeck *et al.*, 2002). One interesting finding is that within the neural assembly representing a prepared sequence, we can distinguish components which correspond to each individual operation in the sequence—and moreover, that these components are active in parallel in the prefrontal planning representation, even though the operations they correspond to occur sequentially. I will discuss the significance of this later in the paper.

A Minimalist model of *The man grabs a cup*

I turn now to the syntactic representation of a cup-grabbing sentence. I will adopt a version of Minimalism (Chomsky 1995) as my syntactic framework. As already mentioned, Minimalism is interesting in that it allows us to express the hypothesis that sentences reporting the same episode in different languages have the same 'underlying' syntactic structure (termed 'logical form' or LF). But it is also interesting in proposing that LF representations have a very simple recursive form. In Minimalism, LF structures are made up of copies of a single template, called an **X-bar schema**, which is illustrated in Figure 2(a). The notion of an X-bar schema

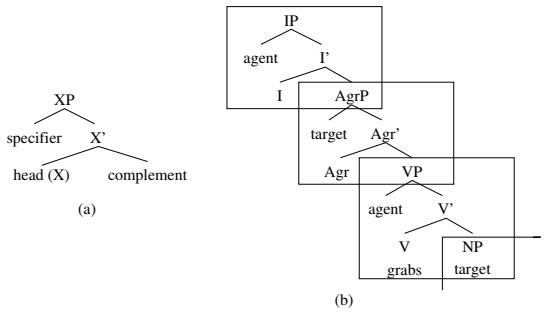

Figure 2: (a) The XP schema. (b) LF of *The man grabs a cup*

is lexicalist in inspiration: the basic idea is that the syntactic structure of a sentence is formed from sub-components, each of which is contributed by one of the lexical items in the sentence. The idea in X-bar theory is that lexical items all contribute the same type of structure no matter what class they belong to: the X-bar schema is a representation of this basic structure. In each case, the lexical item itself occupies the **head** of the structure. The structure also contains two other positions, the **specifier** and the **complement**, at which the semantic arguments of the head can appear. These positions can be occupied recursively by other XPs realising these

arguments, and in this way complex syntactic structures can be defined.

The notion of X-bar schemas is in fact used in many linguistic formalisms, but in Minimalism, where these schemas are used to represent underlying syntactic structures, they have wider application than normal: it is not just lexical items which contribute XPs, but also *inflections* on these items. Figure 2(b) shows the LF of the sentence *The man grabs a cup*. This structure is made up of four instances of the XP schema (each identified by a box). The two lower XPs are familiar: VP is contributed by the verb, and the complement NP of this verb is contributed by the noun denoting the target object. The two higher XPs, IP and AgrP, are associated with inflections on the verb: the head of IP is associated with an inflection agreeing with the subject, and that of AgrP is associated with an inflection agreeing with the object. In English, there is sometimes an overt verb inflection agreeing with the subject (for instance the inflection *-s* in *grabs*). There are no object agreement inflections. However, there are such inflections in other languages: since LF is assumed to be invariant over translations, the projections associated with these inflections appear in the LF of the English sentence, even though they have no overt phonological content.

Although I have drawn the LF representation as a static structure, Minimalists express many syntactic principles by referring to the the process of creating, or **generating** LF representations. This process is not intended to model the actual process by which humans produce sentences: rather, it is a formal device to allow the description of an infinite set of sentences. In Minimalism, many syntactic principles are expressed as constraints on the process of building LF structures. Some of these principles require that elements of an LF structure *move* from one position to another while it is being created. Movement operations serve two different purposes within the theory. Firstly, they support an account of variation between languages: the idea is that an element which moves from one position to another in an LF structure can be appear at in either position at surface (PF) structure, and different languages can have different conventions about where it appears. Secondly, movement operations enable statements of syntactic dependencies between positions in an LF structure. For instance, in English there is a syntactic relationship between a verb and its subject: as just mentioned, the verb must agree with its subject. Since the subject and verb can be quite distant from one another in a syntactic structure, we must describe how this relationship is established; in Minimalism, the account makes reference to movement operations.

There are various types of movement in Minimalism: I will discuss two of these. The first is **NP movement**. In Minimalism, the subject and object originally appear inside the VP, where the verb assigns their thematic roles ('agent' and 'patient' in our example). But they must move to higher positions to be assigned something else called 'Case': the subject must move to the specifier of IP to be assigned 'nominative' Case by the head of IP (I), and the object must move to the specifier of AgrP to be assigned 'accusative' Case (by Agr). One motivation for NP movement is to create two positions for both the subject and the object, to support an account of languages with different word ordering conventions: subjects appear 'early' some languages (e.g. English) and 'late' in others (e.g. Māori), and there is similar variation in object position. Another motivation has to do with syntactic agreement phenomena. To discuss this, I must introduce the second type of movement, **verb raising**. The verb of a sentence originally appears at the head of VP, but is required to raise successively to the heads of AgrP and IP to 'check' its inflections at these positions. (In fact, it is the verb's 'agreement features' which are checked, rather than its overt inflections, so it must raise even if it has no overt morphology.) The verb's object agreement features are checked at Agr, the head which assigns case to the object NP, and its subject features are checked at I, the head which assigns case to the subject NP. Verb raising is motivated partly because it creates some alternative possible word orders, allowing the verb to appear 'early' (as in French or Māori) or 'late' (as in English). But it also allows a simple account of agreement: in IP and VP, the subject and object come into exactly the same local configuration with the verb, so the two types of agreement are explicitly modelled as instances of the same phenomenon.

A sensorimotor interpretation of LF

There are some formal similarities between the sensorimotor and syntactic models sketched above. One similarity is that both models propose a notion of 'basic building blocks'. The sensorimotor model draws on Ballard *et al.*'s account of deictic routines, which makes the strong claim that all sensorimotor processing consists of recursively structured sequences of deictic operations. The Minimalist model makes a similarly strong claim: that all LF representations consist of recursively structured applications of the X-bar schema. Even though our task is to relate the LF of a particular sentence to a particular piece of sensorimotor processing, it is interesting to try and express this relationship as a manifestation of a more general relationship, between the basic building blocks of the two types of structure: i.e. between X-bar schemas and deictic operations. Obviously, it would be very nice if this general relationship existed. But there is also a good methodological reason for looking for a general relationship: any proposal about a link between syntax and the sensorimotor system should make testable predictions, which go beyond the evidence on which it is based. Accordingly, I will begin by proposing a general hypothesis: that for any sentence S reporting a concrete episode E, *each X-bar schema in the LF of S describes exactly one of the deictic operations involved in experiencing E*.

The proposed correspondence between X-bar schemas and deictic operations is shown in Figure 3(a). Recall that a deictic operation takes place in an initial context and brings about a new context, creating a reafferent sensory state as a side-effect. I propose that each component of the X-bar schema

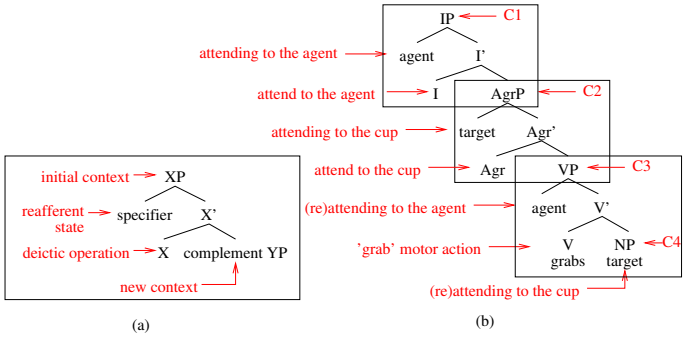

Figure 3: (a) Sensorimotor interpretation of a single XP schema. (b) Sensorimotor interpretation of the LF of *The man grabbed a cup*.

denotes one aspect of this process: XP denotes the initial context, X denotes the operation itself, the specifier denotes the reafferent state and the complement denotes the new context. Note that as a corollary of this definition, a right-branching sequence of XPs describes a sequence of deictic operations, with each operation bringing about the initial context of the next operation. LF representations are largely right branching, so the definition is consistent with the idea that LF structures represent deictic routines.

I now return to the cup-grasping episode. As summarised in Figure 1, experiencing this episode involves a sequence of four deictic operations. The LF of *The man grabs a cup* is a right-branching sequence of four XPs, so the sensorimotor interpretation of XPs just proposed certainly makes the two structures line up. The correspondences predicted by the interpretation of XPs are as follows: IP describes an action of attention to the agent (and its sensory consequence), AgrP describes an action of attention to the target (and its sensory consequence), VP describes the action monitoring routine associated with the 'grasp' motor programme (and its sensory consequence), and the NP complement of VP describes the endpoint of the action, in which the agent is haptically attending to (i.e. holding) the target. The sensorimotor interpretation is shown in detail in Figure 3(b).

Some parts of this interpretation are intuitively plausible. For instance, the idea that a V denotes an action monitoring routine seems a sensible way of characterising it in sensorimotor terms. And if V denotes a particular monitoring routine, then it is plausible that the specifier of V is a position where the subject is assigned the 'agent' thematic role: as suggested in the sensorimotor model, when we are monitoring an action, we evoke a dynamic representation of the agent as a reafferent consequence. Finally, it is plausible that the complement of V denotes the state in which the agent is holding the target. For one thing, this is the state which the action monitoring process brings about, as required by the sensorimotor characterisation of complement position. But in addition, the state of holding the target is a state in which an object representation is evoked—a 'haptic' representation of the target object. This accords well with the fact that the occupier of the complement position is an NP: at a first approximation it is reasonable that NPs denote object representations. There are other parts of the interpretation which need to be considered more carefully. I will discuss two of these.

One interesting feature of the interpretation is that IP and AgrP denote attentional actions. IP and AgrP have some specific roles in the Minimalist model of LF: how does their sensorimotor characterisation accord with the syntactic roles they have to play? To begin with, note that it makes sense that the specifiers of IP and AgrP should be NPs. If I and Agr denote actions of attention to objects, then we expect their specifiers to be the object representations which result—and as just noted, NPs can reasonably be thought of as contributing object representations. But more importantly, characterising IP and AgrP as attentional actions allows us to say something about the role these projections play in the wider sentence. In the Minimalist model, IP and AgrP provide the positions to which the subject and object NPs must raise 'to get Case'. If IP and AgrP are attentional actions, we can give a neat sensorimotor interpretation of this requirement. In the sensorimotor model, as already discussed, it is necessary to establish the agent and patient attentionally before we can monitor an action involving these individuals. Perhaps the requirement that 'NPs raise to get Case' in fact derives from a much more basic constraint on sensorimotor processing that has nothing to do with language at all: that objects have to be attentionally established before they can participate in cognitive routines. This interpretation of Case assignment gives considerable substance to IP and AgrP—projections whose syntactic motivation is often questioned by non-Chomskyan linguists.

I turn now to verb raising. In Minimalism, the inflected verb appears initially at V, but then raises to Agr and I: it can be pronounced either 'early' (at I) or 'late' (at V). If an LF structure describes a sensorimotor sequence, then verb movement involves things being pronounced *out of sequence*: if a verb is pronounced at I, then the motor action is being reported too early, before it actually occurs, while if it is pronounced at V, then the attentional actions denoted by its inflections are being reported too late, some time after they actually occurred. Is there any sensorimotor correlate of this loss of sequential information? I suggest there is, but it requires a slight amendment of the interpretation of an X-bar schema proposed above.

The revised interpretation draws on the idea mentioned briefly above, that episodes are stored in working memory as prepared sequences of sensorimotor operations. The amended definition assumes that an LF representation describes a sensorimotor sequence not as it is actually experienced, but as it is *replayed from working memory*. The idea that semantic representations are simulated experiences already has quite wide currency in 'embodied' models of semantics; see e.g. Gallese and Goldman, 1998; Feldman and Narayanan, 2004). In my model, which already envisages a strong sequential structure to sensorimotor processing, 'simulating an experience' can be interpreted quite concretely: it

involves rehearsing a particular sequence of sensorimotor operations stored in prefrontal cortex. Now recall that prefrontal representations of prepared sequences represent all their component operations *in parallel*. In my amended sensorimotor interpretation of an XP schema, each XP denotes a 'replayed' sensorimotor operation. The specifier still denotes the reafferent state resulting from this replayed operation, and the complement still denotes the new context it brings about (which is now a 'memory context' rather than an actual one). But crucially, heads now report sensorimotor operations indirectly: rather than reporting replayed sensorimotor operations themselves, heads report the planning representation in prefrontal cortex which *enables* their replay. Since all of the operations in a prepared seqence are active in parallel in the planning representation, heads can report all the prepared operations at once, and therefore the verb and its agreement inflections can appear at any head position in the LF structure. To summarise: we can give an interesting sensorimotor interpretation of NP movement, drawing on the structure of deictic routines, and an interesting sensorimotor interpretation of verb raising and agreement inflections, drawing on a model of how experiences are stored in and replayed from working memory.

Towards a model of language processing and language learning

The interpretation of LF just proposed is a very radical one. It largely dispenses with the Minimalist account of how LF structures are 'generated', instead expressing constraints on LF structures in terms of constraints on sensorimotor routines, and on the form of working memory representations. However, this reinterpretation has some advantages: in particular, it opens the way for a model of sentence *processing* which makes reference to the Minimalist notion of LF. As well as providing a detailed sensorimotor interpretation of LF, my forthcoming book (Knott, in press) presents a neural network model of sentence generation. In this model, producing a sentence simply involves rehearsing a sensorimotor sequence in working memory, in a special mode where sensorimotor signals can have overt linguistic side-effects. Since the signals representing the agent, patient and action each occur multiple times in a rehearsed sequence, infants have to learn which signals should result in overtly spoken words in their native language. The network model thus provides an implementation of 'parameter-setting' which should be recognisable by Minimalists, even though it is expressed as a processing model. At the same time, the network is also able to learn rich representations of the surface structure of its exposure language, of the kind which are emphasised in usage-based models of grammar: it can learn a variety of idiomatic constructions, as well as general syntactic parameters.

In my book I conclude that a sensorimotor interpretation of LF is not only promising as the basis for a strongly embodied account of language, but also as the basis for a model of syntax combining insights from Chomskyan and usage-based accounts of syntax, which are normally seen as alternatives to one another. In this paper I have only sketched the arguments for this conclusion, but I hope that interested readers will refer to the book to assess the arguments in more detail.

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.

Averbeck, B., Chafee, M., Crowe, D., & Georgopoulos, A. (2002). Parallel processing of serial movements in prefrontal cortex. *PNAS*, *99*(20), 13172–13177.

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

Barone, P., & Joseph, J.-P. (1989). Prefrontal cortex and spatial sequencing in macaque monkey. *Experimental Brain Research*, *78*, 447–464.

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

Dominey, P., Hoen, M., & Inui, T. (2006). A neurolinguistic model of grammatical construction processing. *Journal of Cognitive Neuroscience*, *18*(12), 2088–2107.

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., & Goldman, A. (1998). Mirror neurons and the simulation theory of mind-reading. *Trends in Cognitive Sciences*, *2*(12), 493–501.

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.

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

Knott, A. (in press). *Sensorimotor cognition and natural language syntax*. Cambridge, MA: MIT Press. (Currently available at www.cs.otago.ac.nz/staffpriv/alik/publications.html)

Novick, J., Trueswell, J., & Thomson-Schill, S. (2005). Cognitive control and parsing: Reexamining the role of Broca's area in sentence comprehension. *Cognitive, Affective and Behavioural Neuroscience*, *5*(3), 263–281.

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

Rizzolatti, G., Fogassi, L., & Gallese, V. (2000). Cortical mechanisms subserving object grasping and action recognition: A new view on the cortical motor functions. In M. Gazzaniga (Ed.), *The new cognitive neurosciences* (pp. 539–552). MIT Press.

Webb, A., Knott, A., & MacAskill, M. (2010). Eye movements during transitive action observation have sequential structure. *Acta Psychologica*, *133*, 51–56.