Syntactic Structures as Descriptions of Sensorimotor Processes

Alistair Knott

In this paper I propose a hypothesis linking elements of a model of theoretical syntax with neural mechanisms in the domain of sensorimotor processing. The syntactic framework I adopt to express this linking hypothesis is Chomsky's Minimalism: I propose that the language-independent Logical Form (LF) of a sentence reporting a concrete episode in the world can be interpreted as a detailed description of the sensorimotor processes involved in apprehending that episode. The hypothesis is motivated by a detailed study of one particular episode, in which an agent grasps a target object. There are striking similarities between the LF structure of transitive sentences describing this episode and the structure of the sensorimotor processes through which it is apprehended by an observer. The neural interpretation of Minimalist LF structure allows it to incorporate insights from empiricist accounts of syntax, relating to sentence processing and to the learning of syntactic constructions.

Keywords: embodied cognition; generative syntax; neural models of language; reaching-to-grasp; sensorimotor processing

1 Two Strategies for Investigating the Neural Correlates of Syntax

At least since Chomsky's *Aspects* (Chomsky, 1965), it has been traditional for linguists to think of syntactic models as describing something in the domain of psychology, rather than just as characterising sets of well-formed sentences. In practice, theoretical linguists rarely try to express the detailed claims of their models in terms which would be intelligible to neuroscientists or experimental psychologists. However, if a syntactic theory really does describe the cognitive system, then ultimately we might expect the theoretical devices it employs (representing syntactic structures or operations) to correspond to brain mechanisms which are identifable in their own right. For instance, if a theory uses a particular device to represent how phrases are formed, or how long-distance syntactic dependencies are sanctioned, we might expect to be able to find a mechanism in the brain which 'implements' the device, and ultimately explains why it has a role to play in the overall theory. The mechanism should be independently identifiable using the techniques available to neuroscience. Neuroscience is maturing rapidly; there are now many well-established methods for investigating brain structures and processes at several levels of hierarchy. So eventually we might expect to be able to

I am grateful to the Marsden Fund of New Zealand for support of this work. I would like to thank the two anonymous reviewers for their helpful comments.

ISSN 1450–3417

express a hypothesis linking terminology or devices from syntactic theory to identifiable brain mechanisms.

Poeppel and Embick (2005) provide a useful discussion of this expectation. They argue that it is not a foregone conclusion: that some syntactic theories are expressed at a level of detail which is unlikely ever to correspond to measurable observations in neuroscience. In order to move towards a linking hypothesis, they suggest that syntactic theories should be formulated in ways that highlight the general *classes of computation* which must be carried out by the syntactic system—and to seek theories in which the proposed types of computation are plausibly implementable by neural machinery. It is not clear what syntactic formalisms Poeppel and Embick intend to rule out through this proposal, but it is certainly helpful for a theory to be explicit about the general types of computation which the syntactic mechanism must be capable of, and to focus the search for a linking hypothesis on a search for neural mechanisms which can implement computations with the right properties.

There is another strategy for seeking brain mechanisms underlying syntactic devices, which is in some ways complementary to that proposed by Poeppel and Embick: namely to focus the search for a linking hypothesis on the semantic domains which have been most intensively studied within neuroscience. Promising areas of neuroscience to consider are those which study perceptual and motor mechanisms (which I shall refer to jointly as 'sensorimotor' mechanisms). Models of sensorimotor mechanisms are among the most detailed in neuroscience, because it is through these mechanisms that the brain interfaces with the world, and they are therefore the easiest to study empirically. If we want to look for neural mechanisms underlying a particular syntactic structure, it may be helpful to begin by studying examples of the structure featuring concepts with obvious links to sensorimotor mechanisms, rather than concepts taken from some arbitrary semantic domain. For instance, if we are interested in the neural mechanisms underlying predication, it might be helpful to look first at 'concrete' sentences like *The cat is white* rather than arbitrary predications like *The idea is popular* or *The company is solvent*. We know quite a lot about the neural mechanisms involved in perceiving concrete objects and properties, so if we begin with *The cat is white*, we have a point to start from on the neuroscience side.

The idea of focussing on concrete language is central to a recent programme of research in psychology and neuroscience, investigating the claim that language somehow supervenes on, or recruits, sensorimotor mechanisms (see e.g. Rizzolatti and Arbib, 1998; Glenberg and Kaschak, 2002; Zwaan and Taylor; 2006; Pulvermüller, 2010; Meteyard *et al.*, 2012). These claims form part of a broader 'embodied' conception of cognition which is currently quite influential in cognitive science (Barsalou, 2008). Of course a hypothesis about the neural mechanisms underlying syntax must eventually extend beyond concrete sentences to all sentences. If we look first at concrete sentences, we must remember that we are looking in the sensorimotor domain for *instantiations* of general neural mechanisms, which we expect to also find in other more abstract domains. But it may still be helpful to start by studying concrete sensorimotor domains.

2 A Project Which Combines the Two Strategies

In this paper, I will summarise a research project which pursues both of the strategies outlined above. The aim of the project is to suggest how the devices of a formal syntactic theory can be linked to neural mechanisms. In line with the second strategy just discussed, the project focusses on sentences in a concrete domain, which express propositions that could conceivably be apprehended through well-studied sensorimotor processes. In line with the first strategy, the syntactic formalism used to analyse these sentences is one which emphasises simple computations, for which there is some prospect of a neural implementation.

The project focusses on a single example episode: one in which a man reaches for and grasps a cup. The sensorimotor mechanisms involved in experiencing this episode have been particularly closely studied, so it is possible to formulate a fairly detailed model of the processing that takes place when it is apprehended. On the syntactic side, the project focusses on the sentences which most directly report this example episode: transitive sentences such as *The man grabs the cup*, *He grabs it*, *L'homme prend la tasse*, *L'homme le prend* and so on.[1] My aim is to juxtapose a detailed model of the sensorimotor processes needed to apprehend the cup-grabbing episode against a model of the syntax of the example sentences, and look for formal similarities between the two models. If there are nontrivial similarities, this might provide some insight about how syntax reflects sensorimotor mechanisms, and ultimately, how syntax supervenes on neural mechanisms more generally.

The project is described at length in a recently published book (Knott, 2012). The book introduces both the sensorimotor and syntactic models 'from scratch', since there are few readers who are familiar with the details of both fields: as a result the key ideas are interspersed with a large amount of motivating background material. The purpose of the current paper is to present the key ideas in the book with a minimum of background material. I will begin in Section 3 by introducing the syntactic model, which is a version of Chomsky's Minimalism. In Section 4 I will introduce the overall form of the sensorimotor model I adopt, and in Section 5 I describe the form of the correspondence between syntax and sensorimotor processing which I envisage. In Sections 6 and 7 I introduce a detailed model of the sensorimotor processing through which a cup-grabbing episode is apprehended and stored. In Section 8 I make some proposals about how the syntax of the example sentences can be linked to the detailed sensorimotor model. In Section 9 I introduce a computational model of language processing and language acquisition which draws on these proposals. In Section 10 I consider how the proposals might extend beyond the example cup-grabbing sentences, and note some areas for further work.

3 A Syntactic Model of Transitive Cup-Grabbing Sentences

3.1 Choosing a Syntactic Formalism

To begin with, we need a syntactic formalism in which to describe the structure of the example cup-grabbing sentences. Of course, there are many formalisms to

[1] I will briefly discuss ergative languages and passive constructions in Section 10.

choose from. In the project I am describing, I use an early version of Chomsky's Minimalism (Chomsky, 1995b). In fact I will eventually adopt a heavily revised version of Minimalism which incorporates ideas from more empiricist syntactic frameworks, so I encourage non-Minimalist linguists not to stop reading here! However, I do want to adopt some of the key tenets of a standard Minimalist model. There are several reasons for this, which I will discuss in turn.

To begin with, Minimalism strongly emphasises simple general-purpose computations in the way advocated by Poeppel and Embick. The Minimalist programme aims to reduce to a minimum the amount of theoretical machinery required to generate the well-formed sentences in a language. Two simple operations, Merge and Move (or latterly Copy), do much of the work involved in generating a syntactic structure. I don't want to claim that Minimalism is the only formalism which adopts a small repertoire of basic computational operations. There are other formalisms which place an equal emphasis on computation, and posit an equally minimal repertoire of computations. For instance in categorial grammar, syntactic derivations are produced (largely) by two operations, 'functional composition' and 'type-raising'. And there is interesting research exploring the neural basis of these operations in non-linguistic domains; see for instance Steedman (2002). Nonetheless, Minimalism arguably meets Poeppel and Embick's criteria for linguistic formalisms: if we find neural mechanisms which plausibly implement Merge and Copy, we can make a substantive claim about how syntactic analyses refer to neural mechanisms.

A second reason for adopting Minimalism is that it allows strong claims about linguistic universals to be made. In Minimalism, there are two levels of syntactic analysis for a sentence: **phonetic form** (**PF**) represents its surface word order (among other things), and **logical form** (**LF**) represents the structure which the language processing mechanism delivers to a nonlinguistic semantic/conceptual system. The surface word order of sentences is obviously very different in different languages. One of the interesting claims in Minimalism is that there is a level of syntactic representation, namely LF, where many of these differences disappear, and where generalisations across languages are manifest. For instance, while sentences describing the cup-grabbing episode in different languages have very different PF structures, the claim in Minimalism is that they have the same structure, or at least very similar structures, at LF. Among current syntactic theories, this claim is unique to Minimalism. It is interesting, because it allows for particularly strong statements about the neural basis of syntactic structures: we can claim that an LF structure describes some neural process or representation which directly interfaces with language, and which is the same for all speakers. For instance, consider our project of studying the relation between syntax and sensorimotor processing in the cup-grabbing scenario. It is presumably uncontroversial that people the world over use the same sensorimotor mechanisms to apprehend a cup-grabbing episode regardless of the language they speak. Within Minimalism, we can frame a very strong hypothesis: that the LF of a cup-grabbing sentence (in any language) directly describes or recapitulates the sensorimotor processes involved in experiencing the episode it reports. It is only within a framework like Minimalism that we can express the idea that syntactic representations directly encode sensorimotor processes. Several theorists developing 'embodied' models of language have argued that entertaining the meaning of a concrete sentence involves a process of active sensorimotor simulation (see e.g. Gallese and Goldman, 1998; Jeannerod, 2001; Grèzes and Decety, 2001; Barsalou *et al.*, 2003; Feldman and Narayanan, 2004). In this paper I want to provide a syntactic framework for expressing this idea. But for the moment, my point is just that the Minimalist device of LF allows particularly strong, optimistic statements connecting syntactic structures and language-independent neural processes.

Some further reasons for adopting a Minimalist framework relate to the actual structure of LF as proposed within Minimalist models. I will briefly introduce this structure in the next section, and then describe why it has some appeal as the basis for an account of the neural mechanisms underlying syntax.

3.2 The LF Structure of a Transitive Cup-Grabbing Sentence

Minimalist LF representations have a strongly recursive structure. The primitive structural unit in LF representations is the **X-bar schema**, which is illustrated in Figure 1. X-bar theory (Chomsky, 1970; Jackendoff, 1977) long predates Minimalism,

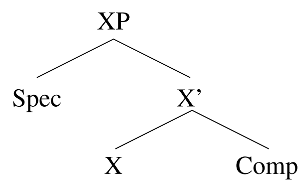


Figure 1: The X-bar schema

and forms an important part of many non-Chomskyan models of syntax, in particular HPSG (Pollard and Sag, 1994). The theory was originally introduced as a refinement of the general hypothesis that syntactic structure is lexical in origin, i.e. that the lexical items in a sentence contribute or 'project' their own elements of local syntactic structure. X-bar theory states this idea parsimoniously, positing that lexical items of all grammatical classes project a phrase with the same basic form, namely the XP structure shown in Figure 1. XP contains a position for the lexical item projecting the structure, called the head (X), and also two positions where semantic material required by the head can appear, the specifier (Spec) and the complement (Comp). These latter two positions can be recursively filled by other XPs.

In Chomskyan models, X-bar syntax has been extended in several directions. In current Minimalism, the X-bar schema is derived from applications of a still more basic structural operation, Merge, already mentioned in Section 3.1. For most of this paper I will retain the terminology of X-bar theory, though I will briefly discuss Merge in Section 10.1. The extension of X-bar theory I will focus on is the notion of **functional projections**, that features heavily in early versions of Minimalism, and is retained in modern Minimalism. Functional projections are XPs headed by non-lexical elements. In Minimalist analyses, the LF of a clause typically contains many functional projections as well as regular lexical projections, resulting in structures containing many more XPs than in other formalisms. These XPs attach to one another predominantly through adjunction to complement positions, creating what could be termed 'right-branching' LF structures, or more accurately, 'complement-branching' structures (since LF encodes hierarchical semantic structure rather than left-to-right linear order). These complement-branching structures, in conjunction with principles allowing movement of syntactic elements from one position to another, support a distinctively Chomskyan style of syntactic model.

The basic LF structure of the transitive sentence *The man grabs a cup* in the Minimalist model of Chomsky (1995b) is shown in Figure 2. As can be seen, the LF

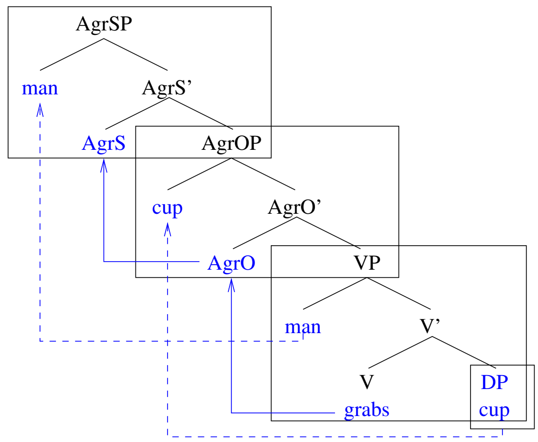


Figure 2: Basic LF structure of *The man grabs a cup*, in the Minimalist model of Chomsky (1995b). Each XP (including the complement of VP) is shown in a box. Head-raising operations are shown with solid lines; DP-raising operations are shown with dashed lines.

structure of a clause is strongly complement-branching. I have omitted the internal structure of noun phrases (which I will call 'determiner phrases' or 'DPs', following Abney, 1987), which introduce some measure of left-branching structure when they appear in specifier positions. But within a DP, Minimalist analyses again envisage a largely complement-branching structure of XPs (see e.g. Abney, 1987; Zamparelli, 2000; Alexiadou *et al.*, 2007). The idea that the LF structure of clauses and DPs is predominantly complement-branching is characteristic of Minimalism.

The most distinctive, and controversial, elements in the Minimalist model of a transitive clause are the functional projections, which are headed by non-lexical material. In Figure 2 there are two functional projections, AgrSP and AgrOP. Functional projections were actually introduced in the precursor to Minimalism, GB (Chomsky, 1981). A novel idea in GB was that XPs can be headed by morphological inflections as well as by whole words: for instance agreement inflections on verbs (and later, tense inflections) were assumed to introduce their own XPs, occupying specific positions in a clause. In Minimalism the idea of functional projections associated with inflections is retained, though these XPs are now headed by the semantic features signalled by inflections rather than by inflections themselves. AgrSP and AgrOP are headed by the main verb's 'subject agreement features' and 'object agreement features' respectively (in this case these are both third person singular). In fact in current Minimalism the LF of a clause standardly includes several further projections that do not feature in Figure 2, CP, TP, vP and several others: for most of this paper I will assume the simplified structure in Figure 2, but I will discuss some of these additional projections in Section 10.2.

A final adaptation of X-bar theory in Minimalism is that X-bar schemas impose a fixed temporal order on the surface form of sentences: the specifier of an XP appears to the left of its complement at PF (Kayne, 1994). For this reason I will continue to use the term 'right-branching' to describe the predominant complement-recursive structure in LF.

The LF structure shown in Figure 2 is generated through a process called **derivation**, which involves the step-by-step merging of XPs, interspersed with various movement operations, latterly reconstrued as Copy operations.[2] The key movement operations are shown in colour in the figure. During derivation, the inflected verb initially appears at the head of VP, but must raise to the heads of AgrOP and AgrSP in turn to 'check' its inflections. The verb's arguments (a subject and an object) initially appear at the specifier and complement of V, where they are assigned thematic roles (AGENT and PATIENT), but they must raise to the specifiers of AgrOP and AgrSP respectively, to be assigned 'Case' by the heads of these projections. At the end of the derivation, we have an LF structure holding multiple copies of the inflected verb, subject and object at different positions.

For Minimalists, the requirements that inflected verbs 'check' their features at higher heads, and that DPs are assigned 'Case' at higher specifier positions, are simply formal stipulations, which are mainly justified by the roles they play in the expression of a larger theory of syntax which economically accounts for a sizeable body of linguistic data. My aim in the current paper is to consider what these devices might correspond to in a model of neural mechanisms. But I will briefly outline the theoretical roles they play in the Minimalist model.

The LF structure produced by derivation illustrated in Figure 2 forms the basis for two accounts. One is an account of syntactic variations between languages—in particular of differences in the canonical ordering of constituents (e.g. subject, verb and object). The other is an account of nonlocal syntactic dependencies within a clause within any given language. The account of cross-linguistic differences turns on the idea that PF structures are formed by 'reading out' LF structures, with the constraint that only one copy of each moved constituent is pronounced. Differences in canonical constituent order between languages are modelled as differences in the conventions governing this read-out process: thus speakers of English (which has subject-verb-object or SVO order) learn to pronounce the subject 'high', and the verb and object 'low', while speakers of Māori (which has VSO order) learn to pronounce the verb 'high'. The Minimalist account of nonlocal dependencies explains agreement relationships between apparently distant syntactic items by appealing to unpronounced material at LF. For instance, in English, there is agreement between subjects and verbs, even though subjects appear outside the local domain of the verb at PF. In Minimalism agreement is explained by referring to the structure of LF, in which the verb is present at the head of AgrSP as well as at the head of VP.

---

[2] I will continue to refer to them as movement operations, though I will certainly adopt the copy theory of movement.

3.3 Prospects for a Sensorimotor Interpretation of Minimalist LF Structures

To summarise the previous section: Minimalist LF structures are formed from instances of the X-bar schema (including many functional projections), joined together into predominantly right-branching structures. They contain two kinds of nonlocality: one resulting from movement (copying) of DPs to Case-assigning positions, and one resulting from successive movement (copying) of Vs to feature-checking positions.

Both these properties of LF structures could potentially have interesting parts to play in an account of the relation between sensorimotor processing and syntactic structure. Firstly, LF structures are created from simple building blocks (XPs) combined in a simple general way. There are some sensorimotor models which argue that sensorimotor processing is likewise decomposed into simple building blocks, recursively combined to create right-branching structures. I will discuss one of these in Section 4. If any relationship can be found between the building blocks of LF and the building blocks of sensorimotor processing, this would be a very interesting discovery. Secondly, there is interesting evidence for multiple occurrences of certain representations during sensorimotor processing (see Section 6.3 for details). For instance, in the process of apprehending an episode in which a man grabs a cup, multiple distinct sensorimotor representations of both the agent (the man) and the patient (the cup) are activated. Perhaps these provide some basis for the Minimalist device of movements occurring during derivation of an LF structure. I will pursue this idea in Section 8.

4 A Proposal about the General Form of Sensorimotor Processing

An interesting proposal about the 'building blocks' of sensorimotor processing was made by Dana Ballard and colleagues (Ballard *et al.*, 1997), in the context of a general model of embodied cognition. In this section I will introduce Ballard *et al.*'s model.

4.1 Deictic Representations

Ballard *et al.* argue that at a certain timescale, which they term the 'embodiment level', cognitive processing is intimately connected to sensorimotor routines interfacing with the physical world. At this timescale, they suggest, cognitive processing engages in a special way with the moment-to-moment deployment of sensory and motor apparatus. The illustrations they provide mainly relate to the role of saccadic eye movements in cognition. Human agents make around 3 saccades per second throughout their waking lives. Each saccade an agent executes results in a very transitory fixation on some visual stimulus, and in a similarly transitory pattern of activity in the agent's visually-derived neural pathways. For instance, the activity in the 'ventral' visual pathway leading to inferotemporal cortex changes dramatically after every saccade (Freedman *et al.*, 2003); so does at least some of the activity in the 'dorsal' visual pathway through posterior parietal cortex (Colby and Goldberg, 1999). Ballard *et al.* call the cognitive representations which reflect the momentary deployment of an agent's sensorimotor apparatus **deictic representations**.

Ballard *et al.* argue that deictic representations play an important role in structuring cognitive processing at the embodiment level. For instance, deictic representations reflecting the current position of the eye can help plan the next saccade: when this is executed, it will generate a new set of deictic representations, so deictic representations and saccades are naturally organised into alternating sequences. To take another example, deictic representations play an important role in the organisation of motor actions. When we want to act on a target object, we typically make a saccade to the object first (Land and Furneaux, 1997; Johansson *et al.*, 2001). This creates deictic representations of the object in visual cortex and in parietal and premotor cortex (Geyer *et al.*, 2000, Murata *et al.*, 2000). Representations in the latter areas encode the object's motor affordances, and are used to select and eventually control a motor action (see e.g. Cisek and Kalaska, 2010). Note that motor actions are not only initiated by deictic representations—they also bring about new deictic representations: for instance, when an agent touches a target object, the tactile sensors of his hand are 'momentarily deployed' to the cup (see e.g. Goodwin and Wheat, 2004), in much the same way that saccades momentarily deploy the fovea to a particular point in the visual field.

4.2 *Deictic Operations*

Ballard *et al.*'s model is interesting because it identifies commonalities between attentional actions (e.g. saccades) and motor actions (e.g. reach movements). Actions of both types are modelled as **deictic operations**: cognitive operations which bring about updates in the agent's physical relationship with the environment, and also in his internal cognitive representations. When talking about deictic operations in general terms, Ballard *et al.* make use of a notion of **context**, which includes information about the agent's cognitive representations at any given time and also about the physical state of the agent and his environment at that time. To formalise their definition a little: a deictic operation is executed in an **initial context**, and results in the establishment of a **new context**; it also generates a **reafferent sensory signal** as a side-effect. The reafferent signal is a deictic representation which provides sensory feedback that the operation actually occurred. For instance, if the operation involves attending to a particular object in the environment, the natural reafferent signal confirming the operation has taken place is a sensory representation of the object. The general form of a deictic operation is shown in Figure 3.

Initial context	Deictic operation	Reafferent signal	New context
C_1	O_1	S_1	C_2

Figure 3: General form of a deictic operation

4.3 *Deictic Routines*

Ballard *et al.* propose that cognitive processing at the embodiment level is organised into sequences of deictic operations called **deictic routines**. A simple deictic routine associated with reaching-to-grasp is shown in Figure 4. This particular routine involves two **cycles**; each row of the table describes one of these. This example

Initial context	Deictic operation	Reafferent signal	New context
C_1	*attend_to_cup*	*cup*	C_2
C_2	*grab*		C_3

Figure 4: A simple deictic routine with two cycles

illustrates the recursive nature of deictic routines: the new context created by the first operation *attend_to_cup* (C_2) is also the initial context for the second operation *grab*.

Note that we need to talk about recursion rather than simple iteration. A point which Ballard *et al.* repeatedly stress is that deictic operations and representations typically cannot be interpreted by themselves: they tend to 'implicitly refer' to the operations and representations which preceded them. For instance, the deictic operation *grab* does not specify a target: the target is specified implicitly by the deictic representation in place at the time when *grab* is executed. According to this model, an agent selects the target of a motor action simply by attending to it, and then activating the general motor programme 'grab'. This motor programme has various free parameters: the location to reach for, the hand shape to form, and so on. The prior action of attention activates deictic representations in motor pathways which fix the values of these parameters. Thinking of the 'grab' programme as a deictic representation accurately reflects the neural mechanisms involved in actually controlling an action. To take another example, the deictic representation *cup* only provides partial information about a cup: it specifies its visual properties, but not its location in the world, since it is always centred on the retina, at the fovea. To establish where the fovea is currently directed, we must refer back to the deictic operation which positioned the fovea, namely *attend_to_cup*.

In the general case, to recover the meaning of a deictic representation, we must make reference not just to the immediately preceding deictic operation, but to some sequence of preceding deictic operations. For instance, in order to interpret the deictic representations at the end of a grab action, when the agent's hand is touching the cup (at context C_3 in Figure 4) we must refer back to the *grab* operation which directly led to these sensations, but interpreting the *grab* operation in turn requires reference back to the preceding *attend_to_cup* operation, as discussed above. The interpretation of deictic operations and representations is right-recursive: it requires recursive reference to a preceding deictic routine.

In Section 6 I will outline a detailed model of reaching-to-grasp which assumes the framework of deictic routines just sketched. But first I will introduce the basic proposal I want to make about the relation between syntax and sensorimotor processing, which is also expressed using the terminology of deictic routines.

5 A Sensorimotor Interpretation of LF Structures and of the X-Bar Schema

As outlined in Section 3, Chomsky's Minimalist model posits that sentences have an LF structure as well as a surface phonetic form: LF structure is relatively invariant over translations, and is composed of instances of the X-bar schema, connected in a largely right-branching way. As outlined in Section 4, Ballard *et al.* propose that sensorimotor processes are organised into deictic routines, which are composed of deictic operations in a similarly right-branching way. Having introduced a model of syntax and a model of sensorimotor processing, I will now formulate some proposals which link these two models. Recall that my proposals are only about the syntax of 'concrete sentences', which describe episodes that can be directly apprehended through sensorimotor processing.

I will first make a general proposal linking syntactic structures to deictic routines.

Proposal 1 The LF of a concrete sentence reporting an episode E provides a description of the deictic routine through which E is experienced.

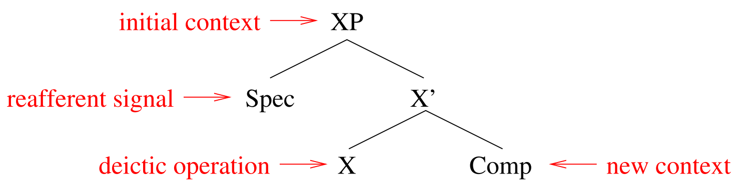
One way of thinking about Propoal 1 is as an expression of the kind of simulationist account of sentence meaning advocated by theorists like Gallese and Goldman (1998) and Feldman and Narayanan (2004). In Minimalism, the LF of a sentence represents its meaning—or at least, as much of its meaning as can be encoded syntactically. The novel thing about the proposal is that it expresses a simulationist account of meaning in a way which links recursively structured syntactic representations to recursively structured sensorimotor routines. This opens the way for a much stronger statement of the proposal, which links the basic *building blocks* of LF structures to the basic *building blocks* of sensorimotor routines. Ultimately, what I want to propose is a general sensorimotor characterisation of the X-bar schema from which LF structures are formed, as follows:

Proposal 2 Each X-bar schema in the LF of a concrete sentence describes a single cycle in the deictic routine described by the LF structure.

More specifically, I want to suggest that each constituent within the X-bar schema has a well-defined sensorimotor interpretation.

Proposal 3 The components of an X-bar schema describe a cycle of a deictic routine, as follows:

- the maximal projection (XP) describes the initial context in which a deictic operation occurs;
- the head (X) describes the operation itself;
- the specifier describes the reafferent signal of the operation;
- the complement describes the new context which the operation brings about. As shown in this figure:



Proposal 2 and its extension Proposal 3 express very strong claims, as they link the basic structural element from which syntactic structures are formed to the basic element from which sensorimotor processes are formed. (Recall we are still restricting our attention to concrete sentences.) One of the attractive aspects of Minimalism is that it envisages simple building blocks for syntactic structures. One of the attractions of Ballard *et al.*'s model is that it envisages simple building blocks for sensorimotor process (at least at the embodiment timescale). I want to suggest that there is a relationship between the building blocks in each domain.

An important idea implicit in the above proposals is that the right-branching organisation of LF structures mirrors the right-branching structure of deictic routines. A corollary of Proposal 3 is that a right-branching structure of XPs describes a sequence of deictic operations, i.e. a deictic routine. This is illustrated in Figure 5, which shows a right-branching structure of two X-bar schemas, XP and YP. The

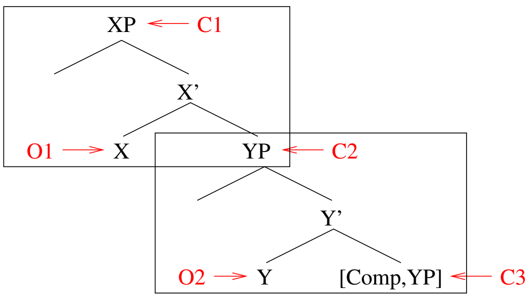


Figure 5: Sensorimotor interpretation of a right-branching structure of two X-bar schemas

red labels show some of the sensorimotor interpretations which are sanctioned by Proposal 3. XP describes the initial context ($C1$) for some deictic operation ($O1$). The complement of XP describes the new context ($C2$) brought about by this operation. Since YP appears at this position, it also describes $C2$, which is therefore also the initial context for a second deictic operation ($O2$), bringing about a third context ($C3$). Thus by Proposal 3, a right-branching strucure of X-bar schemas describes a deictic routine. (Note that if LF structures are in general right-branching, Proposal 1, that LF structures describe deictic routines, is also a corollary of Proposal 3.)

Because Proposal 3 is expressed in terms of the building blocks of LF structures, it makes very strong and specific predictions about the relationship between syntax and sensorimotor processing for *any* concrete sentence. The proposal predicts that we can take any sentence describing a concrete episode, and find a mapping of the right kind between the LF of this sentence and the structure of the sensorimotor processing through which the described episode is apprehended. Given that the accounts of LF structure and of sensorimotor processing are derived from completely separate data (well-formedness judgements vs data about neural processes), using completely different methodologies (syntactic argumentation vs experimental neuroscience), finding a mapping of the predicted kind would provide good empirical support for the generalisation expressed in Proposal 3.

Of course, there is more to LF than the right-branching structure illustrated in Figure 5. As already discussed in Section 3.2, the LF of a transitive sentence is regularly right-branching, but also contains various types of re-entrancy, reflecting movement operations: DP raising to Case-assigning positions and head-raising for semantic feature checking (see the illustration in Figure 2). If the LF structure of a concrete sentence really does describe sensorimotor processing, we should make several further predictions, namely that these movement operations can also be given sensible sensorimtotor interpretations. In the next three sections I will examine these predictions. In Sections 6 and 7 I will introduce a model of the sensorimotor processing required to apprehend a cup-grabbing episode and to store it in working memory. In Section 8 I will examine the relationship between this model and the LF model introduced in Section 3.2, considering both the overall right-branching structure of LF and the re-entrant structures associated with DP movement and head raising.

6 Sensorimotor Processing Involved in Apprehending a Cup-Grabbing Episode

In this section I will argue that the sensorimotor processing through which a cup-grabbing episode is apprehended takes the form of a deictic routine with three cycles, with the structure shown in Figure 6. The argument is given in much more detail in Knott (2012); I will just summarise it here.

Initial context	Deictic operation	Reafferent signal	New context
C_1	*attend_man*	*man*	C_2
C_2	*attend_cup*	*cup*	C_3
C_3	*grab*	*man*	C_4 / *cup*

Figure 6: Deictic routine involved in experiencing a cup-grabbing episode

6.1 Cycle 1: Attention to the Agent

The deictic routine shown in Figure 6 is somewhat more complex than the short routine used for illustrative purposes in Section 4.3. The main reason for this is that it provides an account of the apprehension of a whole *episode*, rather than just of the execution of a motor action. A cup-grabbing episode has an agent, as well as a motor action and a target. I will begin by arguing that whenever an observer apprehends an episode involving an agent through sensorimotor means, the agent must be attended to first.

As a starting point, note that the agent of the grab action—the man—could be someone external to the observer, but could also be the observer himself. In a sentence reporting the observed episode, this difference is reflected in the sentence's subject, and in the verb's subject agreement features (*I grab* vs *He grabs*). Interestingly, apart from these differences, the LF of the sentence is the same in these two scenarios. Proposals 1–3 make a strong prediction from this observation: that the sensorimotor processes through which an agent apprehends his own actions are largely the same as those through which he apprehends the actions of others.

There is certainly support for this general idea in neuroscience: it forms the basis for the well-known 'mirror system hypothesis' (see e.g. Gallese *et al.*, 1996). The hypothesis originated with the discovery of neurons in area F5 of the premotor cortex of macaque monkeys which respond to specific grasp movements whether they are initiated by the monkey itself or by an observed third party (di Pellegrino *et al.*, 1992). It has since been corroborated in many other studies finding common representations for observed and executed actions (see Fadiga *et al.*, 2005 for a summary). The hypothesis has been put to work in several ways in neural models of language, in particular in accounts of language evolution trading on the suggestion that area F5 is the macaque homologue of Broca's area (see e.g. Rizzolatti and Arbib, 1998; Arbib, 2005). But the main use I want to make of the hypothesis is in an account of the very first sensorimotor operation an observer must perform when apprehending an episode.

If we accept that agents use assemblies in premotor cortex to represent the actions of others as well as their own actions, this raises the question of how an agent is able to distinguish between his own actions and those of other people. There are several models of how this distinction is made. My model derives mainly from accounts by Haggard (2008) and Farrer and Frith (2002). The basic idea is that the neural circuitry activating action representations in premotor cortex must be *configured* for either action execution or action observation before representations in this area can be interpreted. There is good general evidence that the brain can switch between alternative general modes of connectivity, implemented by distinct large-scale neural networks (see e.g. Bressler and Menon, 2010). I argue that action execution and action observation require the establishment of different modes of connectivity within the mirror system. In 'execution mode', premotor representations are activated through a well-studied sensorimotor pathway which maps visual representations onto motor affordances (as discussed in Section 4), and in turn they generate overt motor actions. In 'perception mode', premotor representations are activated through a completely different neural pathway through the superior temporal sulcus, specialised for analysing biological movements (see e.g. Jellema *et al.*, 2000), and a mechanism is in place to prevent these premotor representations resulting in overt actions, so that the observer does not reflexively imitate the actions he sees.

The above account can be neatly expressed using the terminology of deictic routines. Action representations in premotor cortex are a variety of deictic representation: they cannot be interpreted in isolation, because in isolation we do not know whether to attribute them to ourselves or to some external agent. In order to interpret them, we must make reference to a prior mode-setting operation. The mode-setting operation can therefore be modelled as a deictic operation.

It is also useful to think of the deictic operation establishing action execution or action perception mode as an action of attention, to the *agent* of the action which will presently be represented in premotor cortex. The operation establishing execution mode is like an action of attention to onesself, since it is this operation which allows an agent to attribute the action subsequently evoked in premotor cortex to himself. There is good evidence that part of an agent's 'concept of self' is implemented in the neural mechanisms which initiate volitional actions; see e.g. Haggard, 2008; Chambon *et al.*, 2012. And there are good grounds for allowing for the possibility of actions of attention to onesself; see e.g. Damasio, 1999 and Critchley *et al.* (2003) for accounts of a 'pre-attentional' representation of the self from which higher-level representations of self can be selectively activated. The operation establishing perception mode is also plausibly attentional in origin. In the most obvious scenario it is triggered by attention to an external agent in the world, who presents a salient enough stimulus to the observing agent to cause him to engage his action perception circuitry. In this case, the deictic operation required to interpret activity in premotor cortex involves a regular action of visual attention *as well as* a mode-setting operation. The reafferent side-effect of this attentional action is a representation of the particular external agent who is attended to.

To summarise: when an observer apprehends an episode in which an agent executes an action, he must attend to the agent (generating a representation of the agent as a reafferent side-effect) before activating a representation of the action in premotor cortex. The agent attended to could be the observer himself or it could be an external agent, but in either case we can speak of an action of attention to this agent, resulting in activation of a representation of this agent.

6.2 Cycle 2: Attention to the Target

If the observed action involves reaching for a target object, the observer must also attend to this object before activating a representation of the action. As already noted in Section 4, an agent executing a reach action typically fixates the target object very early in the reach, probably before a detailed grasp motor programme has been activated (Land and Furneaux, 1997; Johansson *et al.*, 2001). So if the observer of the action is the agent, there is quite good evidence that the target must be attended to before a motor programme is activated. Interestingly if the observer is watching an external agent perform a grasp action, there is also evidence that attention to the target precedes activation of a motor representation. Observers watching an external agent reach for a target typically saccade to the target well before the agent's hand reaches it (Flanagan and Johansson, 2003; Webb *et al.*, 2010). They are able to infer the observed agent's intention, and then put themselves into the same attentional state as the agent while the action is still under way. This establishment of joint attention plays a crucial role in computational models of learning in the mirror system, because it allows agents to use visual representations of their own actions to learn how to recognise the actions of external agents (see e.g. Oztop and Arbib, 2002; Oztop *et al.*, 2005).

To summarise: in a deictic model of the perception of a reach-to-grasp action, two deictic operations must occur before the grasp action itself can be activated. First, the agent of the action must be attended to. This operation is either an action of attention to onesself, if one is executing the action, or an action of attention to an external agent. Second, the target of the action must be attended to. If one is executing the action, this operation is needed to fix the parameters of the motor action to be executed. If one is watching an external action, the operation is executed in order to establish the same attentional state as the observed agent. In either case, it is only once the agent and patient have both been attended to that a single motor programme can be selected.

6.3 Cycle 3: Action Monitoring and Completion

When the observer attends to the intended target object, a set of alternative motor programmes is activated in premotor cortex (the object's 'motor affordances') and these compete amongst one another until a winning programme is selected, either as the programme to be executed (see e.g. Fagg and Arbib, 1998) or as the winning hypothesis about the type of action being performed (Oztop and Arbib, 2002).

Once a motor programme has been selected in this fashion, the character of sensorimotor processing changes. In the case of action execution, a physical action is initiated, which is shaped in real time by the representations currently active in the agent's sensorimotor pathways (Cisek and Kalaska, 2010). The agent's neural pathways, motor effectors and physical environment together implement a dynamical system, which (if all goes well) moves towards an attractor state where the agent is holding the target. In the case of action perception, processing also has the character of a dynamical system, this time a simulated one (see again Oztop and Arbib, 2002; Oztop *et al.*, 2005).

Again it is useful to cast this processing within the framework of deictic routines. Activating the 'grab' representation in premotor cortex is an operation which initiates a dynamical system. This operation has its own reafferent sensory consequences, and also eventually results in a new context in which new deictic operations are possible. I will consider these two aspects of the operation in turn.

What is the reafferent sensory signal associated with the execution of or observation of an action? When we are in the process of executing an action, we are aware of our body moving: there is a characteristic 'match' betwen the pattern of outgoing motor signals and the pattern of incoming sensory signals, because the sensory signals are produced by our motor movements. This match signal appears to be involved in producing the sense of agency that we feel when we perform an action, through which we are able to attribute the action to ourselves (Farrer and Frith, 2002; Farrer *et al.*, 2008). Of course, our conception of self is highly multimodal. Different elements of this construct are activated at different stages of action execution. We have already seen that one component of an agent's conception of self is activated at an early stage during action preparation (Haggard, 2008, Chambon *et al.*, 2012). Another component is activated when an action as actually under way (Farrer *et al.*, 2008). (Chambon *et al.* refer to these two conceptions as 'propsective' and 'retrospective' concepts of self.) The circuits which control action execution must include some mechanism for binding together these different facets of our concept of self.

A very similar point can be made about action perception. As argued in Section 6.1, an observer initially attends to an external agent as a salient object in the world. But when the observer begins to monitor that agent's actions, the agent is represented differently, as a characteristic pattern of movement, rather than as a static object. (See Ramsey and Hamilton, 2010 for evidence that action observation involves dissociable representations of the observed agent as a token individual and as an animate entity.) Again, our conception of an observed agent is a multimodal construct, and different facets of the construct seem to be accessed at different points during action perception: the action perception mechanism must include a mechanism which integrates these different representations of observed agents.

In sum, I argue that an observer attends to the agent at two distinct points in the course of apprehending a reach-to-grasp action— once at the very start, when action perception or action execution mode is established, and once during the process of actively monitoring the action to completion—and that there must be a mechanism for binding together the different representations of the agent obtained through these two different actions of attention.

I now consider the new context brought about by the grasp motor action: namely the state in which the agent is holding the target object. As already noted in Section 4, this state is well modelled within the framework of deictic routines, because it has both physical and attentional components. Successfully executing a grasp action achieves a substantive physical change in the world: the agent's hand moves to achieve stable contact with the target. But it also deploys the agent's tactile senses to the target object, providing him with information about the target in a new modality. Just as there are two actions of attention to the *agent* in the course of apprehending a reach-to-grasp action, there are quite clearly two actions of attention to the *target*, again in different modalities, and again occurring at different times. The target is first attended to in the visual modality, as part of the process of preparing the grasp action, and it is later attended to in the haptic modality, when the action is complete. Again, the structure of the sensorimotor routine plays a critical in the formation of a cross-modal representations of the target object. At the point when a stable grasp is achieved on the object, the visual representation of the object axiomatically corresponds to one of its motor affordances (see e.g. Oztop *et al.*, 2004).

At this point, I have motivated the arguments for the deictic routine illustrated in Figure 6. In the initial context C_1, the observer attends to the agent (who is himself or an external agent), and as a reafferent consequence receives a representation of the agent (*man*) and establishes a new context C_2. In this context, objects in the agent's perispace compete for attention, and the observer attends to a cup, activating a representation of the cup as a reafferent consequence, and establishes a third context C_3. In this context, the agent activates the motor programme *grab*, and an action is dynamically monitored. While the motor programme is under way, a dynamic representation of the agent is activated as a reafferent consequence, and when the action is complete, we enter a final context C_4, in which a haptic representation of the cup is active.

A final thought: it might perhaps be thought that modelling the experience of a reach-to-grasp action as a deictic routine as illustrated in Figure 6 somehow understates the complexity of the action-monitoring process, which is construed as a single stage in the routine, with the same basic form as a simple attentional action. I certainly acknowledge that action monitoring, whether it involves execution or perception of an ongoing action, has greater complexity than an attentional action. Monitoring a reach-to-grasp action takes far longer than attending to an object, and it is well known that the process involves two largely separate neural pathways, one for reaching and one for grasping: see classically Jeannerod (1996) (and for a review see Knott, 2012). But my suggestion is that this type of compositional structure is not *visible to language*. My proposal is that the syntax of language engages with the discrete, *temporal* structure of a sensorimotor routine. And at this level, the process of monitoring an action may in fact be quite simple. A motor programme controlling the dynamics of the combined hand/arm system (whether actual or simulated) can bring about complex changes in this system without having complex dynamics itself; this is an important fact about dynamical systems. If language interfaces with representations that *control* the dynamics of movements, rather than with dynamical movements themselves, the additional complexity of action monitoring may not be visible in the sensorimotor signals that interface with language.

7 Working Memory Representations of Sensorimotor Routines

One outstanding question from Section 5 concerns what's meant by the assertion that an LF representation 'describes' a sensorimotor routine. Clearly I do not want to propose that sentences *directly* report sensorimotor processes, as these arise in real time. We can execute sensorimotor routines without engaging language, and we can produce concrete sentences which are unrelated to our current sensorimotor environment. The standard assumption in psycholinguistics is that sentences produced by a speaker reflect *representations* of events and states, held in that speaker's working memory (Levelt, 1989). I certainly want to adhere to that idea. I therefore need to provide an account of how the cup-grabbing episode is represented in working memory. Ideally, this should be framed within a more general account of episode representations in working memory. Ultimately, I need to re-express the proposals about LF made in Section 5 in a way which connects to an account of working memory episode representations.

In Sections 7.1 and 7.2 I make a suggestion about how concrete episodes are stored in working memory. In Sections 7.3 and 7.4 I discuss how this mechanism interfaces with language, and in the light of this, give a more precise interpretation of LF.

7.1 A Model of Episode Representations in Working Memory

My account of working memory is based on Alan Baddeley's recent model of working memory for episodes (Baddeley, 2000). Baddeley suggests that there is a working memory medium called the **episodic buffer**, which holds semantic representations of episodes, and which interfaces with the well-known **phonological buffer** (Baddeley and Hitch, 1974) in a way which supports language processing.

Baddeley's main argument for the episodic buffer hinges on the fact that experiencing a concrete episode often takes a significant amount of time. (Apprehending an episode involves monitoring it as it occurs; this may take several seconds, often longer.) In order to store the episode in long-term memory, it must be encoded in the hippocampus, as a preliminary to being consolidated in cortical long-term memory (McClelland *et al.*, 1995). But storing associations between representations in the hippocampus can only be achieved through the mechanism of long-term potentiation (LTP), which requires them to be active within about 100ms of each other (Abraham *et al.*, 2002). Baddeley concludes that experienced episodes must be buffered in a working memory medium, and then replayed to the hippocampus at a speed which allows them to be associated through LTP. Additional evidence for the existence of short-term memory representations of observed episodes is reviewed in Swallow *et al.* (2009).

Baddeley does not speculate much about the format in which episodes are encoded in the episodic buffer, beyond requiring that it supports them being transmitted to the hippocampus. There are many models of how episodes are stored in working memory (see e.g. Shastri, 2001, 2002; Chang *et al.*, 2002; Plate, 2003; van der Velde and de Kamps, 2006). I make a new suggestion, which is based on the assumption that concrete episodes are experienced as deictic routines. My suggestion is that a concrete episode like the cup-grabbing episode is stored in the episodic buffer as a *planned sequence of sensorimotor operations*, i.e. a planned deictic routine (see Takac and Knott, 2013 for an implemented model). This view of working memory episode representations is interesting for several reasons. Firstly, it offers a new solution to the question of how thematic roles are bound to participants in episode representations. Any neural model of episode representation must have a way of identifying the roles played by the different participants in an action (in our case, agent and patient). The deictic routine through which an episode is experienced distinguishes these roles clearly, because they are associated with different serial positions in the routine. Secondly, the neural mechanisms which support the preparation of sensorimotor sequences have been intensively studied, and we know quite a lot about them. If these prepared sequences are examples in a 'concrete' domain of working memory episode representations, then studying their properties may help us formulate a more general model of these representations which extends beyond concrete episodes. Thirdly, viewing working memory episode representations as prepared sequences fits well with their role in Baddeley's model of replay to the hippocampus. The hippocampus is often seen as specialised for storing sequentially structured information (Wallenstein, 1998; Eichenbaum *et al.*, 2004), and is known to support fast replay of sequences (see e.g. Foster and Wilson, 2006; Diba and Buzsáki, 2007). If working memory episode representations are prepared sensorimotor sequences, they can be communicated to the hippocampus by being replayed, in simulation, at high speed, in a mode where each of them activates an assembly in the hippocampus. Finally, thinking of working memory episode representations as supporting simulation accords well with the simulationist accounts of propositional meaning already mentioned (Gallese and Goldman, 1998; Feldman and Narayanan, 2004 and others).

In the remainder of this section I will refine my sensorimotor interpretation of LF in a way which makes reference to the model of working memory for episodes just outlined. The basic idea will be that an LF representation describes a deictic routine *as it is replayed from episodic working memory*, rather than as it occurs in real time. In Section 7.2 I discuss the mechanics of the working memory replay operation, and make a suggestion about the pattern of sensorimotor signals activated during a replayed sensorimotor routine. In Sections 7.3 and 7.4 I discuss the linguistic reflexes of these signals, and give a more precise sensorimotor definition of LF which makes reference to these.

7.2 *Sensorimotor Signals Active During a Replayed Sensorimotor Routine*

As just mentioned, we know a lot about the neural mechanisms which store prepared sequences of sensorimotor operations in working memory. These mechanisms are mainly in dorsolateral prefrontal cortex (see e.g. Barone and Joseph, 1989; Averbeck *et al.*, 2002; Averbeck and Lee, 2007) and the supplementary motor areas (Shima and Tanji, 2000).

One interesting property of prefrontal sequence representations is that while they support the execution of sensorimotor operations in sequence, they actually identify the different prepared operations individually, and *in parallel*. For instance, within the prefrontal representation encoding the cup-grabbing routine, it is possible to identify assemblies encoding each of the three prepared operations, *attend_man*, *attend_cup* and *grab* (see especially Averbeck *et al.*, 2002). I will make this explicit by designating the prefrontal sequence plan $plan_{attend\_man/attend\_cup/grab}$.

There is also evidence that planning representations remain tonically active while the planned sequence is being executed (see Averbeck and Lee, 2007; also the computational models of Rhodes *et al.*, 2004 and Takac and Knott, 2013). When a cup-grabbing episode is replayed from working memory, we therefore generate a mixture of sustained and transient signals to be activated, as shown in Figure 7. The sustained signals are all active in prefrontal areas; the transient ones occur in

Sustained signals	Transient signals			
	Initial context	Deictic operation	Reafferent signal	New context
$plan_{attend\_man/attend\_cup/grab}$	c_1	*attend_man*	*man*	c_2
$plan_{attend\_man/attend\_cup/grab}$	c_2	*attend_cup*	*cup*	c_3
$plan_{attend\_man/attend\_cup/grab}$	c_3	*grab*	*man*	c_4 / *cup*

Figure 7: The time course of signals occurring during the replay of the cup-grabbing episode in working memory

the sensorimotor areas which are active during actual sensorimotor experience, in accordance with simulationist models.

Another important property of the prefrontal sequence-planning mechanism is that it makes use of a representation of 'the current context' which is updated after each operation. Executing a planned sequence of operations relies on tonically active representations of the planned operations, but also on a dynamically changing representation of context. There are several different models of these context representations (see e.g. Dominey *et al.*, 1995; Houghton and Hartley, 1995; Beiser and Houk, 1998; Takac and Knott, 2013). I will remain neutral about their exact nature; in Figure 7 I just refer to them as c_1–c_4.[3]

7.3 The Interface Between Replayed Sensorimotor Sequences and Language

In Section 7.1 I proposed that an LF representation describes a sensorimotor routine as it is replayed from working memory. Before I consider how this proposal applies to the cup-grabbing sentence, I want to state it as concretely as possible, by providing an explicit proposal about how sensorimotor signals interface with linguistic representations at the level of neural circuits.

[3] Note that in earlier examples of deictic routines I used capitals to denote contexts ($C_1 \ldots C_n$). These were understood as denoting contexts as represented by a theorist watching an agent: they combine objective information about the agent's current physical state with information about his current sensorimotor representations. 'Subjective' contexts ($c_1 \ldots c_n$) are basically an agent's internal representations of objective contexts. In an account of how sensorimotor signals interface with language, we can only make reference to subjective contexts.

I first assume that there is an area of the human brain which encodes motor plans associated with linguistic actions (e.g. planned articulatory actions), which I will call the **premotor output area**. This roughly corresponds to what others have called the 'phonological output buffer', and have localised in parts of Broca's area and adjacent areas of prefrontal and articulatory premotor cortex (see e.g. Henson *et al.*, 2000; Sahin *et al.*, 2009).

I further assume that the human brain contains a number of neural circuits specialised for concrete language which I will call **interface circuits**, each of which links an area expressing sensorimotor signals to the premotor output area, so that activating sensorimotor signals in these areas can activate arbitrary motor plans (in particular articulatory plans). These circuits can hold learned associations between individual sensorimotor concepts and individual motor movements: they allow the production of atomic symbolic gestures, of the kind posited in Bickerton's (1995) account of 'protolanguage'. There may be several interface circuits, linking different classes of sensorimotor signal to the premotor output medium. I assume each of these circuits evolved separately, through an adaptive mutation allowing a particular class of sensorimotor concepts to be associated with overt expressive behaviours.

Now consider what happens when a sensorimotor routine stored in episodic working memory is replayed, and we generate a pattern of sensorimotor signals like the one shown in Figure 7. This will produce a pattern of signals in the premotor output area. Importantly, this pattern of output signals need not reflect the pattern of sensorimotor signals in every detail. There may not be an interface circuit for every sensorimotor area. And different circuits may express sensorimotor signals at different levels of detail. Thus when a prepared deictic routine is replayed, the pattern of signals evoked in the premotor output area will reflect the pattern evoked in sensorimotor areas, *as filtered* by the interface circuits. With this idea in mind, I will state more precisely how I suggest we can interpret the LF of a concrete sentence.

Proposal 4 The LF of a concrete sentence reporting an episode E provides a description of the pattern of signals activated when the deictic routine through which E is experienced is replayed from episodic working memory, including signals activated in the premotor output area.

7.4 *The Replay Mechanism as the Basis for an Account of Head Raising*

Proposal 4 suggests that LF reflects the structure of replayed sensorimotor routines, but also the properties of the interface circuits linking sensorimotor areas to the linguistic output medium. A key question now is: what are these interface circuits, and what are their properties?

As already mentioned, not all sensorimotor areas participating in a replayed deictic routine need have interfaces to linguistic outputs. Informationally speaking, there is a lot of redundancy in the pattern of signals active in a replayed deictic routine. During replay, each deictic operation is expressed in two different media, one encoding the operation as it is planned, the other encoding it as it is experienced (or rather, simulated). For instance, in the pattern shown in Figure 7 there is a representation of the motor operation *grab* in the planning medium (where it is tonically active), and another in the 'deictic operation' medium (where it is transiently active in Cycle 3). Likewise there are multiple representations of *man* and *cup*: these feature both as actions of attention and as transient reafferent object representations.

The presence of *sustained* representations of deictic operations in the pattern of activity created by a replayed deictic routine prompts an interesting suggestion about the neural mechanisms underlying the Minimalist account of head raising. Recall from Section 3.2 that an inflected verb generated at the head of VP must raise to the head of AgrOP and then the head of AgrSP, to check its 'agreement features'. AgrSP and AgrOP are headed by bundles of subject and object agreement features respectively, while VP is headed by the verb. But the verb is allowed to appear at higher head positions, and its agreement features are allowed to appear at lower head positions. The Minimalist account of head-raising allows, indeed requires, heads to appear 'out of position': it is through this device that a Minimalist analysis explains the extended syntactic domain of the verb and its agreement features.

If, as I am suggesting, a right-branching LF structure describes a temporal sequence of deictic operations, and if the head of each XP in an LF structure signals a deictic operation (see Proposal 3), then the principles which sanction head-raising must be interpreted as allowing heads to describe deictic operations *out of sequence*. They allow the head of an XP to report not just the deictic operation presently occurring, but *all* deictic operations in the deictic routine currently being rehearsed, including those which have already occurred and those which have yet to occur. A natural way of explaining this is to suggest that the linguistic signals of deictic operations come from the area where they are planned, and are therefore tonically active in parallel during replay, rather than from the areas where they are transiently expressed one by one. I will suggest a further elaboration of the 'sensorimotor' interpretation of an XP to this effect:

Proposal 5 The head (X) of an XP in the LF structure of a concrete sentence describes a deictic operation in a replayed deictic routine *as it is represented in the prefrontal area storing the deictic routine*.

Given that the prefrontal area holding prepared deictic routines represents their component deictic operations tonically, and in parallel, it follows from Proposal 5 that all the heads in a right-branching structure of XPs describe the same set of deictic operations.

To summarise: the suggestion made in Section 7.3, that an LF structure describes a *replayed* sequence of deictic operations (Proposal 4), recommends itself not only as a way of incorporating reference to working memory representations in a sensorimotor characterisation of LF. It also opens the way for a promising account of the neural basis of head-raising, which is grounded in known facts about the representation of prepared sequences in prefrontal cortex. If LF describes a temporal sequence of operations, then the Minimalist device of head-raising within LF can be understood as a way of encoding the presence of tonically active representations within the neural areas from which surface language forms can be produced. And we know that there are such representations when a prepared sensorimotor sequence is replayed.

Independently of these considerations, there is also good evidence that the processing of verbs and their inflections (the elements involved in head-raising) does indeed involve the dorsolateral prefrontal cortex. Left dorsolateral prefrontal cortex is an area associated with the production and interpretation of verbs (see Perani *et al.*, 1999; Tranel *et al.*, 2001; Pulvermüller *et al.*, 1999) and the processing of verb inflections (see Shapiro and Caramazza, 2003; Shapiro *et al.*, 2012).

8 A Sensorimotor Interpretation of the LF of the Cup-Grabbing Sentence

We can now consider how the general proposals just made about the LF structure of concrete sentences apply to our particular cup-grabbing scenario. We have a Minimalist model of the LF of a transitive sentence reporting a cup-grabbing episode (see Figure 2). And we have a model of the pattern of neural signals activated when the deictic routine through which a cup-grabbing episode is experienced is replayed from episodic working memory (see Figure 7). Proposal 4 asserts that the LF of the cup-grabbing sentence can be construed as a description of the pattern of signals activated during replay—i.e. that the LF structure in Figure 2 can be thought of as a description of the pattern shown in Figure 7. Proposal 2 asserts that in addition, each X-bar schema in the LF structure describes one cycle of the replayed deictic routine. And Proposal 3, extended by Proposal 5, suggests roles for each constituent in each X-bar schema. In this section I will examine these assertions, and in the process make some further general suggestions about how to interpret the Minimalist account of LF in sensorimotor terms.

8.1 General Alignment of the LF Structure and the Deictic Routine

At the level of X-bar schemas, the LF of a transitive cup-grabbing sentence aligns well with the deictic routine it is supposed to represent. The LF structure contains three XPs, and the deictic routine contains three cycles. We can therefore interpret each XP as describing a single cycle of the deictic routine. The predicted sensorimotor interpretations are shown in Figure 8. AgrSP describes Cycle 1 of the routine (an action of attention to the agent), AgrOP describes Cycle 2 (an action of attention to the patient), and VP describes Cycle 3 (execution/monitoring of the 'grab' motor programme). According to Proposal 3, each X-bar schema describes a context-update operation: XP describes the context in which a deictic operation is executed and its complement describes the new context which it brings about. This allows us to interpret the right-branching chain of XPs (AgrSP, AgrOP, VP, DP) as describing the four contexts c_1–c_4 in the deictic routine. The general form of the LF structure certainly corresponds well to the general form of the deictic routine.

In Sections 8.2 and 8.3 I will consider the interpretation of specific elements within the LF structure.

8.2 Sensorimotor Interpretations of Argument Positions

I will first consider the sensorimotor interpretations of the positions in the LF structure of the clause at which the verb's arguments appear. There are four of these.

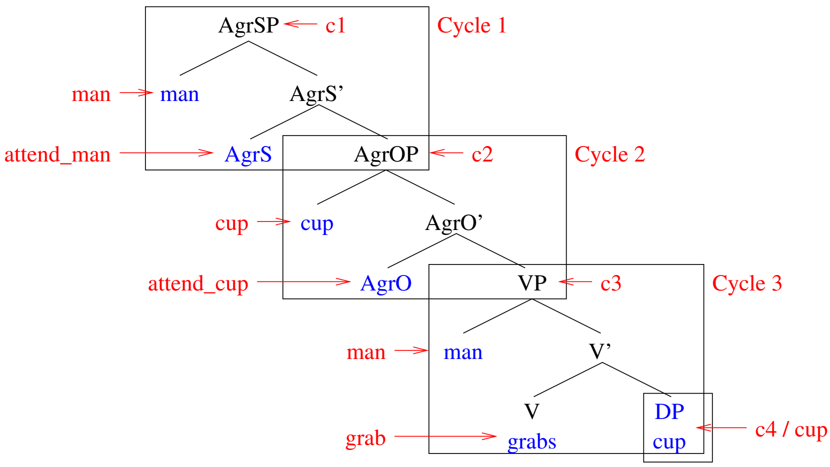


Figure 8: Sensorimotor interpretation of the LF structure of *The man grabs a cup*. Sensorimotor interpretations are shown in red.

The subject *man* appears at the specifier of AgrSP and the specifier of VP. The object *cup* appears at the specifier of AgrOP and the complement of VP.

If we assume the alignment of X-bar schemas and deictic routine cycles shown in Figure 8, and the sensorimotor interpretations of specifier and complement positions made in Proposal 3, these positions correspond exactly to the positions in the deictic routine where representations of *man* and *cup* are active. The specifier of AgrSP is predicted to describe the reafferent sensorimotor signal *man*—and the word *man* appears at this position. The specifier of AgrOP is predicted to describe the reafferent sensorimotor signal *cup*—and the word *cup* appears at this position. The specifier of VP is predicted to describe the second activation of the signal *man*, this time as a reafferent signal generated during monitoring of the motor programme—this is the other position where the word *man* appears. So far so good.

The remaining position, the complement of VP is predicted to describe the consequent state of the action, in which the agent is holding the cup. This is the other position where the word *cup* appears. This position is interesting syntactically, because it is the only place where an argument appears as a complement, rather than as a specifier. Within the sensorimotor model, we can make two predictions about it. On one hand, since the word *cup* occupies this position, we predict it will describe the reafferent signal *cup*. But equally, since it is a complement position, we predict it to describe the consequent state of the action taking place in the third cycle of the deictic routine, namely the *grab* action.

As discussed in Section 6.3, the consequent state of the *grab* action does indeed have a special dual status. This action is a substantive motor action, bringing about a change in the world. But it is also an attentional action, providing information about the cup in the modality of touch. The stable grasp state is *axiomatically* a state in which the cup representation is active, because it is at this point that the function mapping visual object representations onto goal motor states is trained. The dual status of the VP complement as an argument position *and* a complement position perfectly reflects this built-in identification of the target representation with the goal motor state.

In summary, the sensorimotor interpretations predicted for all four argument positions by Proposal 3 are quite plausible. In fact one might well go further, and suggest they may provide helpful insights into the Minimalist concepts of 'specifier' and 'complement'. Recall that the aim of the paper is to seek a linking theory connecting syntactic constructs to neural mechanisms. Without such a theory, syntactic constructs like 'specifier' and 'complement' are justified purely through the roles they play in a larger theory of syntax. Sensorimotor characterisations of these terms provide a way of thinking about them empirically which is entirely independent from the role they play in syntax. If the independent conceptions of 'specifier' and 'complement' line up with the conceptions motivated from syntactic theory, as they seem to in the cup-grabbing example, this has the character of an empirical result: it suggests that the empirical conceptions may be *the basis for* the theoretical conceptions, explaining why these conceptions play a useful role in a theory of syntax.

8.3 Sensorimotor Interpretations of Head Positions: V, AgrS and AgrO

I will next consider the sensorimotor interpretations of head positions which are predicted by Proposal 3. Proposal 3 asserts that the head of an XP denotes a deictic operation: thus the items appearing at V, AgrS and AgrO are predicted to describe the three deictic operations in the deictic routine through which the cup-grabbing episode is apprehended.

The case of V is quite straightforward: the word which appears at this position (*grabs*) is easily thought of as describing the third operation in the deictic routine, activation of the *grab* motor programme. The more interesting cases are AgrS and AgrO. In Minimalism, these heads hold the 'agreement features' of the subject and object respectively. Agreement features classify the subject and object into broad semantic categories along various dimensions, principally PERSON, NUMBER and GENDER, though exactly how categories are defined within these dimensions varies from language to language. The sensorimotor interpretation of LF sees these bundles of agreement features (e.g. 3RD PERSON SINGULAR, MASCULINE) as 'descriptions of attentional operations'. Does this proposal make sense?

To begin with, it should be noted that Agr projections as originally proposed in Minimalism were seen as carrying features with no real semantic content; Agr features were traditionally checked *and then erased*, so they were not visible in the LF structure presented to the semantic system. In my interpretation, Agr heads carry as much semantic conent as other heads; they describe sensorimotor operations, just like other heads: so they are certainly used somewhat nonstandardly in my interpretation. In any case, the question to consider is whether it is plausible to think of Agr features as describing attentional operations.

I will first consider the PERSON feature, and in particular the distinction between 1ST and 3RD person. DPs carrying the 1ST person feature (e.g. *I*) refer to the speaker; those carrying the 3RD person feature (e.g *he*, *she*, *the man*) refer to an external agent. At least in this case, I suggest that agreement features can be very well interpreted as descriptions of attentional operations. In the account of sensorimotor processing given in Section 6, the operation through which an observer attends to the agent of a cup-grabbing episode is also fundamental in implementing that observer's ability to distinguish between himself and external agents. Recall that action episodes are represented within the observer's mirror system. In my account, the circuitry in the mirror system has to be configured for execution of actions or for perception of external actions before signals in the system can be interpreted—and these mode-setting operations must also be construed as attentional operations, which direct attention to the agent of the observed episode. So at least with regard to the 1ST and 3RD PERSON features of AgrS, it makes perfect sense to think of agreement features as describing attentional operations.

Whether this idea extends to other agreement features, and to AgrO as well as AgrS, is a matter for further work. Certainly there is a plausible attentional basis for the distinction between 1ST and 3RD PERSON at AgrO. The attentional action through which an agent establishes himself as a reach target is very different from that which establishes an external target. There are specialised sensorimotor pathways controlling actions directed towards the self, and these actions have different dynamics from those directed externally (see e.g. Petreska and Billard, 2009; Ferri *et al.*, 2010). As regards the other PERSON feature, 2ND PERSON, attending to the addressee is substantially different from attending to a third party. When a speaker is producing an utterance, the addressee must already have been established as a focus of attention, so rehearsing an action of attention to the addressee involves a special kind of reattention which could plausibly be linguistically marked. There is also some reason to suppose that the distinction between SINGULAR and PLURAL is attentional in origin. For instance, the brain area which most plausibly encodes syntactic plurality, the left temporoparietal junction (Domahs *et al.*, 2012) is also activated by attentional operations parsing a visual stimulus as a group rather than as a single entity (Yamaguchi *et al.*, 2000). A computational model of the attentional origin of the singular-plural distinction is given in Walles (2010).

I will conclude with some comments about GENDER features. These features are much more open-ended semantically than PERSON and NUMBER, and much more language dependent (Corbett, 1991). Can these features be thought of as reflecting aspects of an attentional action? I think this is also plausible. The important thing to note is that attentional actions like *attend_man* or *attend_cup* involve top-down establishment of open-class object representations as well as saccades to external points in the world. When an observer executes *attend_cup*, he activates a representation of a cup as a search goal, which can be matched against object representations arriving bottom-up, so there is some way of knowing whether the action is successful (Tomita *et al.*, 1999; Hasegawa *et al.*, 2000; Hamker, 2004). I suggest that while most of the open-class properties of attentional operations are expressed through their reafferent sensory consequences (i.e. through nominal expressions), we can also read *some* of these properties from representations of the operations themselves, which are signalled by heads. Recall my assumption that there are 'interface circuits' linking areas evoking sensorimotor signals to a language-specific premotor output area (see Section 7.3). These circuits are allowed to have different capacities. I suggest that the circuit which generates linguistic reflexes of the open-class semantic properties of deictic operations has rather limited capacity, and that GENDER agreement features are generated through this channel.

In summary, there is some support for the sensorimotor interpretations of the head positions AgrS, AgrO and V. The interpretation of V is certainly plausible. The interpretation of the 1ST and 3RD PERSON features that appear at AgrS and AgrO is also very plausible, and again seems to provide some insight into the neural basis of these agreement features. But more work is needed to determine whether there are sensorimotor correlates of the other agreement features that can appear at AgrS and AgrO.

8.4 A Sensorimotor Interpretation of V-Raising and Agreement

In Section 7.4 I suggested that the heads of XPs should be understood as describing deictic operations 'as they are planned' rather than as they are evoked in sequence. This idea, taken together with known properties of the sequence-planning mechanism in prefrontal cortex, led to an interesting sensorimotor interpretation of the Minimalist device of head-raising. Now that we have sensorimotor interpretations for the AgrS, AgrO and V heads in the cup-grabbing sentence, it is useful to reconsider this sensorimotor account of head-raising, to see how it applies in this particular case.

Consider the English version of the sentence, *The man grabs the cup*. In the surface structure of this sentence, the subject *The man* appears outside the VP, and is therefore syntactically somewhat remote from the verb. Any account of syntax has to explain why the verb's inflection has to agree with the subject, even though the verb is not near the subject. The Minimalist account of head-raising explains this by positing that the inflection *-s* signals an agreement feature which actually 'belongs' at a position above VP, where it *is* near the subject. It is allowed to appear on the verb because at LF the verb 'covertly moves' up to the position where it really belongs.

In my proposed sensorimotor interpretation of head-raising, the subject agreement inflection on the verb is a signal of properties of the attentional action which established attention on the agent of the cup-grabbing action. (Specifically, it signals that this operation involved configuring the mirror system for action observation rather than action execution.) The reason why the inflection is allowed to appear on the verb is that linguistic reflexes of attentional actions are generated from the region where they are planned, and are therefore tonically active. According to this interpretation, the phenomenon of agreement is seen as reflecting the machinery through which episodes are stored in, and replayed from, working memory. The interpretation suggests a specific neural mechanism which accounts for the syntactic phenomenon of subject-verb agreement, as it is accounted for within the Minimalist model.

Most models of syntax include a device allowing agreement features to spread through a syntactic structure. For instance, this is achieved through unification operations in models like HPSG, Tree-Adjoining Grammar and Combinatory Categorial Grammar. Does my proposed sensorimotor interpretation of agreement apply equally well to the account of agreement features given in these other frameworks? I think there are two aspects of the Minimalist account which make it a particularly good vehicle for expressing this sensorimotor interpretation of agreement. Firstly, the Minimalist model envisages head movement taking place at a language-independent level of syntactic representation. The suggestion that deictic operations interface with language through a planning medium where they are tonically active makes no reference to particular languages; the Minimalist device of expressing movement at a language-independent level of syntactic representation is therefore particularly appropriate. Secondly, a Minimalist LF structure can be neatly interpreted as a description of a sequence of operations. It is therefore particularly suitable for expressing an account of agreement phenomena grounded in a neural model of prepared sequences.

8.5 A sensorimotor Interpretation of DP movement, Case and Thematic Roles

There are two kinds of movement at LF: movement of heads to higher head positions, and movement of argument DPs to Case-assigning positions. In this section I will consider what sensorimotor interpretation can be given to DP movement.

The Minimalist account of DP-raising supposes that the argument DPs of a verb initially appear at positions within the VP. In our example, the subject of *grabs* appears at the specifier of VP and its object appears at the complement. At these structural positions, the verb's arguments are assigned thematic roles: the specifier assigns AGENT role and the complement assigns PATIENT role. But Minimalism also requires arguments to be assigned 'Case'. Case can only be assigned by functional heads above VP: the heads AgrS and AgrO assign nominative and accusative Case to their specifiers respectively. So the subject and object DPs must raise to these specifier positions.

As with the other theoretical devices in Minimalism, the principle which requires arguments to raise to Case-assigning positions is justified purely through the formal role it plays in a complete model of syntax which neatly accounts for a large body of linguistic data. The idea that argument DPs must have Case is simply stipulated: there is no proposal about what Case 'is', in the same way as there is no proposal about what specifiers and complements 'are'. (Or rather, it is assumed that the principle 'DPs must raise to Case-assigning positions' corresponds to *some* neural mechanism, but there are no proposals about what this might be.) Does the sensorimotor interpretation of LF allow us to say anything about this principle?

Case Assignment I suggest that there is a very clear sensorimotor interpretation of the principle that DPs must raise to get Case. So far I have argued on several grounds that the functional projections AgrSP and AgrOP describe actions of attention to the agent and the target of the cup-grabbing episode, while the VP projection describes the monitoring of a motor programme. In sensorimotor terms, the requirement that the subject and object appear in AgrSP and AgrOP projections above VP translates as a requirement about the structure of sensorimotor routines—namely that an observer *must attentionally establish* the agent and target of a grab action before he can actively monitor a motor programme involving these individuals. Within a sensorimotor model, this requirement is completely justifiable in its own right. In fact this principle is at the heart of Ballard *et al.*'s conception of deictic routines. In order to monitor a motor programme involving multiple participants, in Ballard *et al.*'s model, an observer must first attend to these participants, one by one, to set up the deictic representations which instantiate the free parameters of the motor programme.

If we are thinking about LF in sensorimotor terms, we can now give a very clear account of the functional projections which assign Case to DPs. These XPs describe the attentional operations which establish the conditions under which the motor programme can be monitored. The general principle that Case is assigned 'by a functional head to its specifier' (which is an important part of the Minimalist account of Case) also has a very clear sensorimotor interpretation. A functional head describes an action of attention, and its specifier describes the deictic representation which this action results in. The deictic representation clearly *depends on* the action of attention. I suggest this dependence is the basis for the Minimalist idea that a specifier depends on, or is licensed by, its head.

Thematic Role Assignment Now that we have a sensorimotor interpretation of the 'higher' subject and object positions above VP, can we find an interpretation of the subject and object positions within VP, the specifier and complement of VP? In Minimalism, these are the positions where the verb's arguments receive 'thematic roles' (namely AGENT and PATIENT). What can we say in sensorimotor terms about these positions? Again, the sensorimotor interpretation of LF seems illuminating. The VP projection describes the cycle of the deictic routine in which the grab action is dynamically monitored. During this cycle, as discussed in Section 6.3, new representations of the agent and the target become active, in new modalities connected with the motor system. While action-monitoring is under way, a representation of the agent as an animate entity is activated as a reafferent side-effect. And when action-monitoring is complete, a representation of the target as a goal motor state becomes active. These points in the routine are described by the specifier and complement of the VP respectively. I suggest that the sensorimotor interpretations of these VP-internal positions explain why they assign the thematic roles they do—and also help us to understand the semantics of thematic roles by showing how they are grounded in sensorimotor representations (in concrete cases such as ours). The AGENT role is assigned by the specifier of VP because this position describes an animate representation of one of the action participants: and the meaning of 'AGENT' in this context comes largely from the nature of this animate representation. The PATIENT role is assigned by the complement of VP because this position describes an affordance-based representation of the other action participant: and the meaning of 'PATIENT' in this context likewise derives largely from the nature of this representation.

DP Movement Finally, we have to seek a sensorimotor interpretation of the Minimalist mechanism by which arguments 'move' from their VP-internal positions to their Case-receiving positions. In the Minimalist model this mechanism is entirely distinct from the mechanism by which verbs raise to higher head positions. We already have a sensorimotor account of verb raising. Can we give an account of DP raising?

Given the ideas suggested in this section, it is clear that the sensorimotor account of DP-raising should relate to the fact that both the agent and target are represented twice, in different modalities, in the course of the deictic routine involved in experiencing the cup-grabbing episode. None of these representations are tonically active during the routine, so their appearance at multiple positions in

LF is certainly due to a mechanism distinct from that underlying head-raising.

My suggestion is that the Minimalist device of DP-raising reflects the associative neural mechanisms through which representations of the agent and target in different modalities are tied together to form multimodal representations. There must be circuitry enforcing certain axiomatic correspondences between representations in different modalities. To create a multimodal agent representation, there must be circuitry linking the reafferent signal activated during action monitoring with the reafferent signal activated by the first action of attention. These two signals axiomatically represent the same object: the agent. To create a multimodal target representation, there must be circuitry linking the reafferent motor state active at the endpoint of the action with the reafferent signal activated by the second action of attention. Again these axiomatically represent the same object: the target. Note that these circuits must link representations purely in virtue of their structural positions in the deictic routine. I will not discuss how they might be implemented neurally, but I do suggest that there must be such circuits to explain how multimodal object representations are learned, and that these circuits provide a plausible neural basis for the Minimalist account of DP raising, which links particular structural positions within LF.

8.6 A Sensorimotor Interpretation of the Generative Mechanism

In Minimalism, LF structures are produced by a generative mechanism. The set of possible LF structures is infinite, so they cannot be enumerated: instead this set is defined indirectly, by characterising the mechanism which creates these structures. I have suggested a sensorimotor interpretation of complete LF structures. But it is also important to have some account of 'the mechanism which produces all possible LF structures'. What might this correspond to in the sensorimotor model?

It is hard to find a direct sensorimotor interpretation for the generative mechanism as it is proposed in Minimalism. The mechanism defined in the Minimalist model proceeds from the bottom up: the lowest projection at LF is created first, and higher XPs are successively adjoined to this. (In our example, the VP would be created first, and then merged successively with AgrOP and AgrSP.) If a right-branching LF structure describes the representation of a temporal sequence, as I suggest, then the Minimalist generative mechanism describes a process whereby this representation is assembled *in reverse*, beginning with the last elements. I cannot see anything in the sensorimotor model which corresponds to this process. The model includes an account of sensorimotor sequences being stored in working memory, but these sequences are stored, and replayed, in the order they are experienced.[4] While there is a good sensorimotor interpretation of the LF structure produced at the end of a derivation, I suggest there is no good interpretation of the generative process understood as a procedural mechanism.

To be clear: I do not want to say that the Minimalist generative mechanism 'does not describe neural processes'. I only want to say that it does not describe neural processes *when understood as a procedural mechanism*. The Mininalist gen-

4 The only evidence I am aware of for *reversed* replay of experienced sequences is in studies of hippocampal representations of spatial location (Foster and Wilson, 2006; Diba and Buzsàki, 2007).

erative mechanism creates representations (LF structures) which I argue describe neural processes in considerable detail. And there are certainly components of the generative mechanism which pick out well-defined features of these processes. To take an example, consider the Minimalist idea that constituents 'move' from lower to higher positions within an LF structure while it is being derived. In one sense, I am saying that there is no such thing as 'movement' of constituents within LF. I do not think there is any neural mechanism corresponding to movement *as such*. But as already argued, I think there are good sensorimotor interpretations of the structures in LF which *result* from movement in the Minimalist account.

Of course, we still need to give a sensorimotor account of 'the mechanism which produces all possible LF structures'. This is a central component of the Minimalist model of grammar. But a sensorimotor account of this mechanism will be quite a radical departure from the Minimalist account. I will conclude this section by considering what this account will look like.

What are the constraints on possible LF structures, if these are thought of in sensorimotor terms? In the sensorimotor interpretation of LF, a right-branching LF structure describes a replayed sequence of sensorimotor operations. If we want to characterise the set of possible LF structures, we must specify *in general* what sequences of sensorimotor operations are possible. We have already seen that there are several general constraints on the sequences of sensorimotor operations an observer can execute. For instance, an observer cannot execute a motor routine without having attended to the participants involved (Section 8.5); an observer cannot attend to a target object before having attended to the agent (Section 6.1). These are the kinds of sequencing constraint which feature heavily in Ballard *et al.*'s model of deictic routines. So part of a sensorimotor account of the generative mechanism will probably involve enumerating constraints resulting from the embodied nature of cognitive processing, of the kind studied by Ballard *et al.* But there are also properties of LF structures which the sensorimotor model sees as reflecting properties of the working memory mechanism which allows an observer to store and replay the deictic routines he experiences (Section 7.4) and properties of the associative mechanisms which support the creation of multimodal object representations (Section 8.5). And finally, there are properties of LF structures which are suggested to reflect the nature of the interface circuits linking sensorimotor areas of the brain to a language output area (Section 7.3).

In summary, the 'sensorimotor' characterisation of the space of possible LF structures will be partly an account of the constraints on the sequential structure of deictic routines, partly an account of the neural mechanisms which store and replay these routines, and which exploit the structure of these routines to learn basic object representations, and partly an account of the neural interfaces between sensorimotor and language areas. Note that the first two parts of this account of LF structures are essentially accounts of embodied sensorimotor cognition: they do not make any reference to specifically linguistic representations or mechanisms at all. The only references to specifically linguistic representations are in the account of interface circuits.

9 A Minimalist-Inspired Model of Language Processing and Language Learning

9.1 The Minimalist Account of the LF-PF Interface

In Minimalism, the surface form of a sentence (its PF) is read from the terminal nodes of its LF structure during derivation, in a process called 'spellout'. The rules governing spellout are language-specific: an infant growing up in a particular language community has to learn a set of rules particular to this language. The rules to be learned relate to the positions at which constituents are pronounced. As discussed in Section 3.2, the LF of our example sentence contains two copies of the agent and patient and three copies of the inflected verb: at PF there is only one copy of each. The Minimalist proposal, in a nutshell, is that 'surface' syntactic differences between languages result from different conventions about which copy of these repeated elements is overtly pronounced.

The Minimalist account of the interface between LF and PF plays two related roles within the overall theory. Firstly, it contributes to a parsimonious model of the differences between languages. These differences are attributed to the mechanism which maps between LF and PF representations. Thus, for instance, we can give an account of languages with different constituent orderings (Subject-Verb-Object versus Verb-Subject-Object and so on) in a way which localises these differences to a single module of the grammar. Secondly, because differences between languages are localised to the LF-PF interface, we can tell a relatively simple story about the learning mechanisms which allow infants to acquire their native language. The mechanism responsible for creating LF structures is assumed to be largely innate. (We are allowed to assume this because LF structures are language-invariant.) In order to learn the syntax of their native language, infants only need to learn how to map LF structures to PF structures. The space of possible hypotheses to test is relatively small and well-structured: the infant just needs to learn the values of a small number of discrete parameters—for instance, whether to pronounce the subject 'high' or 'low'.

9.2 Problems with the Minimalist Account

While the Minimalist account of PF is neat in several respects, there are several well-known problems with it. I will mention three of these.

Firstly, it is hard to square the Minimalist account of PF with an account of sentence processing—that is, with an account of the psychological processes which take place in a speaker producing a sentence, or in a hearer interpreting a sentence. While research into sentence generation and interpretation is still at a fairly early stage, there is good reason to think that both processes are at some level 'incremental'—i.e. that syntactic representations are generated in roughly the order they are produced in (for generation) or heard in (for interpretation). For instance, there is evidence that speakers create representations of early constituents of a sentence first, so that they can begin talking while still in the process of planning later constituents (see e.g. Levelt *et al.*, 1999); likewise, hearers start to generate interpretations of a sentence as soon as the earliest constituents are heard (see e.g. Tanenhaus *et al.*, 1995). Minimalist derivations happen from the bottom up, as discussed in Section 8.6. Since the bottom of an LF structure corresponds to the end of a sentence at PF, and PF structures are read from LF structures, it is hard to see the derivational mechanism *understood as a procedure* as a representation of sentence processing mechanisms. Minimalist theorists frequently assert that Minimalist models describe neural processes (see e.g. Marantz, 2005; Hornstein, 2009), but these assertions tend to be about the general nature of structure-building computations in language rather than about the way these computations are ordered.

Secondly, Minimalism has no account of linguistic structures that are defined in the surface form of sentences. In several syntactic frameworks (see e.g. Fillmore *et al.*, 1988; Goldberg, 1995), descriptions of linguistic structures can include reference to particular surface words as well as to abstract syntactic categories. Some aspects of linguistic structure certainly seem well described in terms of abstract syntactic categories like VP and DP, which are defined recursively and hierarchically. But other aspects seem best modelled simply as patterns involving particular word forms. The clearest examples of such patterns in language are **idioms**: phrases like *how's it going* or *by and large*, which appear to deliver a meaning collectively, rather than individually. The main problem for Minimalism, as forcefully argued by Jackendoff (2002), is that idiomatic linguistic structures appear to interact with abstract grammatical structures in ways which are hard to model if all structure is assumed to be created at LF. For instance, it is sensible to analyse the verb phrase *take X to task* as a grammatical phrase featuring certain specific word forms, in virtue of which it receives a conventionalised meaning ('criticise X'). The phrase conforms to a regular syntactic pattern, and the position *X* can be occupied productively by any DP, but the word stem *take* and the words *to* and *task* cannot be varied productively: the phrase has the meaning it does because it features this pattern of specific words. Minimalist analyses have difficulty modelling partially idiomatic constructions of this kind, especially when the idiomatic elements are discontinuous, as in this example.

Thirdly, the Minimalist model of learning is unlike any other account of learning in current cognitive science. Theories of how humans learn in cognitive science are normally expressed as computational models: for instance neural networks or Bayesian reasoning systems. A central insight gained over the last twenty years or so is that these computational models are very powerful—certainly powerful enough to learn rich representations of surface structures in natural language with very little prior knowledge (see e.g. Pantel and Lin, 2002). A simple type of neural network called a **simple recurrent network** can learn rich representations of the sequential patterns in surface language (Elman, 1990, Christiansen and Chater, 1999). Experiments with computational learning systems lend support to 'empiricist' models of language acquisition, which posit that infants use general-purpose learning mechanisms to acquire syntax, rather than elaborate innate knowledge. The empiricist model of development is supported by evidence that infants' earliest syntactic constructions are defined around particular lexical items, and are therefore idiomatic in nature (see e.g. Lieven *et al.*, 1997; Tomasello, 2003). According to empiricist models, infants learn adult syntax by progressively abstracting away from concrete constructions featuring specific words (see e.g. Tomasello, 2003; Macwhinney, 2005).

To some extent, these difficulties facing the Minimalist model all stem from the way it construes the 'generative mechanism' defining the space of well-formed sentences. The fact that LF structures are generated from the bottom up makes the mechanism unsuitable as the basis for an account of sentence processing. The lack of an account of sentence processing makes it hard to express the Minimalist model of infant syntactic development as a computational mechanism. (Computational models of syntactic development are typically also processing models, which receive their training sentences incrementally, word by word.) The lack of a computational account of learning in Minimalism in turn limits what the theory can say about surface structures in language, since these are best analysed as the product of a learning mechansism.

I have argued that the Minimalist model of LF supports an interesting account of the neural basis of syntactic representations, grounded in an account of sensorimotor processing and working memory. But I have also argued (Section 8.6) that in order to formulate this account, we must abandon the Minimalist model of derivation, because it does not square with the sensorimotor interpretation of LF structures. This opens the way for an account of the relationship between LF and PF structures which is more compatible with models of sentence processing and syntactic development. In the next section, I will introduce a new computational model of sentence processing and syntactic development, whose form is inspired by empiricist models of language processing and language learning, but which also retains the Minimalist conception of LF—interpreted in sensorimotor terms—and the Minimalist idea of parameter-setting.

9.3 An Account of Language Processing and Language Learning

My sensorimotor interpretation of LF puts us in a position to address all three problems described in the previous section. Firstly, and most importantly, it provides an ideal basis for an account of sentence processing. Its central claim is that the LF of a sentence *describes a neural process*: namely the process of replaying an episode representation held in working memory. In this section I will propose that the neural mechanism which implements the described replay process is also the mechanism *through which the sentence is produced*—or at least, a central part of this mechanism. In the paper so far I have argued that thinking of LF as a describing a replayed sensorimotor sequence helps us express Minimalism's essentially declarative account of syntactic structure in a way which makes reference to neural mechanisms. I now want to suggest that a sensorimotor conception of LF has the *additional* advantage of supporting an interesting account of sentence processing.

Until now, the working memory mechanism which allows an experienced sensorimotor sequence to be stored and replayed has not been thought of in relation to language processing at all. The mechanism was introduced in Section 7.1 as part of a model of long-term memory for episodes: it provides a means for buffering experienced episodes so they can subsequently be stored more permanently in the hippocampus (and later still in cortex). It is possible to imagine this whole mechanism predating language altogether. I will begin by sketching an account of language evolution in which the replay mechanism did indeed predate language, and was co-opted by evolution for a new role in communication, supporting the production of word sequences and the learning of syntax. Then I will introduce a neural network model of the circuitry that evolved to co-opt the replay mechanism.

Background Assumption: Two Stages of Language Evolution In Section 7.3 I envisaged a point during human evolution when a collection of 'interface circuits' evolved, allowing agents to produce overt behavioural reflexes of their internal sensorimotor representations (see Section 7.3). These circuits allow agents to learn a vocabulary of atomic behavioural symbols. Several theories see the evolution of such circuits as the first major step in the evolution of human language (see e.g. Bickerton, 1995).[5] In many accounts, these circuits support the production of *sequences* of atomic behaviours, to enable an open-class vocabulary of behavioural symbols, and to permit the production of multi-symbol utterances (see in particular Jackendoff, 1999). I will assume interface circuits support sequential behaviours in this way. However, when interface circuits first evolved, I assume they were not used in any systematic way together with the working-memory episode replay mechanism. For instance, they may originally have permitted behavioural reflexes of an agent's *current* sensorimotor signals, rather than of sensorimotor signals retrieved from working memory. I now suggest that at some later evolutionary point, a second piece of language-related neural circuitry evolved, which allowed agents to produce sequentially structured behavioural signals conveying detailed information about *whole episode representations* rather than just about individual sensorimotor signals. I engisage that this circuitry co-opted the working memory replay mechanism. The replay operation generates a pattern of sensorimotor signals whose serial structure reflects the structure of the episode being replayed. Via the interface circuits, it also generates a sequential pattern of signals in the premotor output medium, as discussed in Section 7.3. These signals still need to be converted into overt motor movements. My suggestion is that the circuit which evolved to co-opt the replay mechanism for a communicative purpose transforms the sequence of signals evoked in the premotor medium during replay into an overt sequence of motor movements.

The Control Network The interface circuits which allow behavioural reflexes of sensorimotor signals generate *premotor* movement signals. These signals do not necessarily result in overt movements, but in general the most active premotor signal will be selected for overt execution (see e.g. Fagg and Arbib, 1998). However, even a strongly activated action signal in premotor cortex can be withheld, if the agent has learned a cognitive control strategy which demands this (see Cohen *et al.*, 2013 for a review). In the model I propose, the network which co-opts the replay mechanism for a communicative purpose learns a control strategy which selects just a *subset* of the premotor signals activated during replay for actual execution. I will call this network the **control network**.

The control network's purpose is to produce behavioural representations of replayed episodes which are short and efficient. Recall that there is considerable redundancy in the signals evoked in the premotor output medium when the cup-grabbing deictic routine is replayed from working memory. There are two activations of a signal reflecting the agent, two activations of a signal reflecting the

[5] There are also alternative theories; see e.g. Wray (1998).

patient and a constant activation of a signal reflecting the whole deictic routine as it is planned. Only one version of each of these signals needs to be expressed behaviourally, provided there is a convention about which versions are expressed. The role of the control network is to learn and implement these conventions.

In Minimalist terms, the control network can be thought of as the device which maps LF structures onto PF structures. The repeated activations of sensorimotor signals during replay of a working memory episode correspond to the multiple copies of subject, object and inflected verb at different positions in LF (see Section 8). In the Minimalist model, infants must learn which copy of each repeated element should be pronounced: this is exactly what is learned by the control network.

If we consider which brain region might plausibly implement the control network, an interesting candidate is Broca's area. There is good evidence that Broca's area and surrounding prefrontal regions play a role in syntactic processing, particularly during sentence generation (see e.g. Bookheimer, 2002). However, they also have a role in implementing cognitive control strategies. In fact, the clearest effect of damage to Broca's area itself is impaired cognitive control: 'an inability to override habitual or prepotent response behaviours' (Novick *et al.*, 2005). The role I envisage for the control network is precisely to override prepotent behavioural signals in premotor cortex, so it is natural to localise it in Broca's area.

A Neural Network Model of Sentence Generation I will now outline a model of the control network, and its interaction with the episode rehearsal mechanism and the interface circuits. The overall model is a neural network of sentence generation. A diagram showing its basic architecture is shown in Figure 9. The model is described in detail in Takac *et al.* (2012).

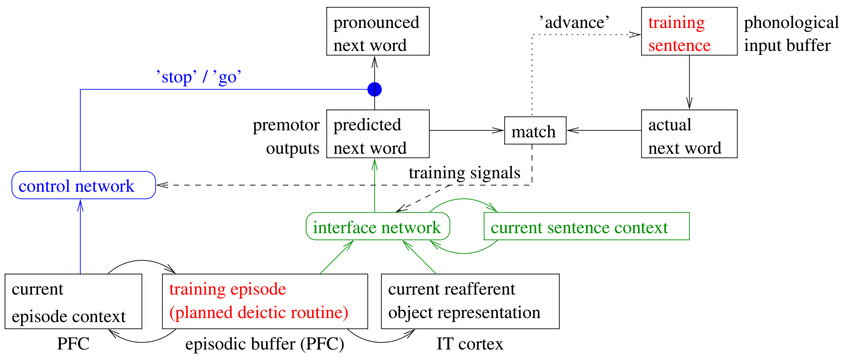


Figure 9: A neural network model of the episode rehearsal system, the interface circuits and the control network

The model is trained on pairs of episodes and sentences, which are shown in red in Figure 9. Each training episode is a representation of a transitive motor action, stored as a planned deictic routine. When replayed, this produces a sequence of episode context representations, and a sequence of reafferent representations of the agent and target of the action (see Section 7.2). Each training sentence is represented as a replayable sequence of words, stored in the phonological input buffer.

Thus episodes and sentences are stored in separate media in working memory, but these media interact, as envisaged by Baddeley (2000).

The interface circuits, here called the 'interface network', are shown in green. The interface network maps sensorimotor signals onto premotor behavioural signals. I will assume these premotor signals are articulatory plans—i.e. word forms, or 'words' for short. The interface network is trained by an error signal ('match' in the figure) which compares the words it predicts from its sensorimotor inputs to words in the training sentence replayed from the phonological input buffer. I assume words in the phonological input medium are also represented as articulatory plans (see e.g. Goldstein and Fowler, 2003) and can therefore be matched to articulatory plans in the output modality.

The interface network can learn to generate a single word from an input sensorimotor signal, but it can also learn to generate a sequence of words from a single sensorimotor input, as envisaged in the models of Bickerton and Jackendoff. This is achieved through the use of a recurrent context representation: the **current sentence context**. A recurrent context representation is used in some form in almost all neural network models of sentence processing. It holds a representation of the sequence of words processed so far, shaped by learning to support prediction of the next word. After each word is generated, the current sentence context is updated to reflect the generated word, and the phonological input buffer advances to the next word, giving the network an opportunity to learn the word which follows the generated word.

The control network, which learns to select a sequence of premotor outputs for overt execution, is shown in blue. This network learns to produce a binary control signal ('stop' or 'go') as a function of the current context representation in the episode rehearsal system—the **current episode context**. This is a representation of 'sensorimotor context' rather than sentence context: it is updated at each cycle of episode rehearsal as discussed in Section 7.2. The control network is also trained by the 'match' signal, but in a different way. If the predicted next word matches the actual next word in a given context, the control network learns to permit words to be overtly pronounced in this context, and causes the input buffer to advance to the next word. If the predicted word does not match, the network learns to refrain from pronouncing words in this context, and does not advance the input buffer. The control network learns language-specific conventions about when to 'read out' words. The conventions it learns depend on the constituent order (SVO, VSO, SOV etc) of the sentences it is trained on.

The full model alternates between two modes of iteration. In one mode, the control network iterates through episode contexts until it reaches a context where a word can be pronounced. The interface network then produces as many words as it can from the current sensorimotor signal, iterating to a new sentence context after each word produced. When it cannot confidently predict the next word, the control network takes over again. The effect of this interaction between the two networks is to combine a fairly traditional empiricist model of sentence processing and sequence learning with a recognisably Minimalist model of parameter setting. If we interpret LF structures as describing replayed deictic routines, then the control network can be interpreted very straightforwardly as a computational model of the mechanism through which infants learn to map LF onto PF representations.

(Note that the control network learns a function which makes no reference at all to the content of words—it only refers to context representations expressing the current state of an episode being replayed. The rules it learns are abstract structural rules, like those proposed in Minimalism.) At the same time, the recurrent component of the interface network allows it to learn rich representations of surface structure in the training data. The sentence context is updated after each word is produced. With training, the interface network can learn to map a single sensorimotor signal onto an idiomatic sequence of words. More interestingly, it also allows the learning of constructions involving a mixture of idioms and productive syntax, for instance 'discontinuous' idioms like *take X to task*. Another interesting point to note is that the model can be configured so that it learns item-specific idiomatic constructions first, and productive syntactic rules later, consistent with empiricist accounts of syntactic development (Lieven *et al.*, 1997; Tomasello, 2003; Macwhinney, 2005). The details of these results, as well as of the architecture and training of the network, can be found in Takac *et al.* (2012).

The model just described is very preliminary; it must of course be scaled up and extended in several directions. My main point in presenting it here is just to emphasise that thinking about LF as a description of a replayed deictic routine is not only helpful in suggesting a neural basis for the representations of syntactic structure proposed within Minimalism: it also allows the Minimalist model of syntactic structure to be stated in a way that is broadly compatible with empiricist accounts of sentence processing and language learning, and of the role of surface structures in language. The network presented here is an example of one such account.

10 Discussion

The aim of this paper is to express a 'linking hypothesis' connecting syntactic representations, as motivated within linguistic theory, to neural mechanisms, as motivated by experiments in psychology and neuroscience. The approach I have taken is to look in detail at a single example sentence, reporting a specific concrete episode, and at the sensorimotor process through which this specific episode is apprehended. My aim is to express a linking hypothesis which connects the *detail* of a sensorimotor model to the *detail* of a syntactic model. At the same time, I have expressed the hypothesis in very general terms, so that it makes predictions which extend beyond this particular example to other concrete sentences. As discussed in Section 5, in the domain of concrete sentences the linking hypothesis assumes that *any* right-branching LF structure describes a sequence of deictic operations. This hypothesis was extended in later sections. I proposed in Section 7.4 that any right-branching LF structure which is a domain for head-raising describes a sequence of deictic operations as it is replayed from working memory storage. And I proposed in Section 8 that any Case-assigning projection describes an attentional action establishing a participant in a sensorimotor routine, and that any instance of DP-raising reflects associative neural mechanisms involved in learning multimodal object representations. In fact, as discussed in Section 1, I am also committed to extending these general proposals beyond the domain of concrete sentences. The point of studying the sensorimotor domain first is simply to develop hypotheses about neural mechanisms in an area where these are relatively easy to study, so that we can later approach other domains with some idea of what we are looking for.

Pursuing these general proposals obviously forms the basis for a whole programme of research at the interface between theoretical syntax and neuroscience. There are many interesting directions to puruse, some of which are discussed in my book (Knott, 2012) and some of which form part of ongoing work. I will not discuss these here, but instead will conclude with some thoughts about the research programme as a whole: how the proposals square with recent developments in Minimalism, how they relate to existing cognitive interpretations of syntax, and how they can be extended beyond concrete sentences.

10.1 A Sensorimotor Interpretation of Merge?

As noted in Section 3.2, in modern Minimalism the X-bar schema is not the primitive recursive building block of LF structure; the structures formerly associated with X-bar schemas are now derived from applications of the more basic operation Merge (Chomsky, 1995a). Merge is an operation that combines two syntactic objects α and β into a single new object, and labels this object with the constituent α, as shown in Figure 10. In this operation, α plays the role of a head in X-bar

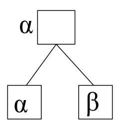


Figure 10: The Merge operation

theory. By applying two successive Merge operations, a structure akin to an XP schema can be created: the complement is joined to the head in the first Merge operation, and the specifier is joined to the result in a second Merge operation. A key difference is that in a Merge-based system, complements and specifiers are not primitives; rather they are defined as positions in structures created by particular combinations of Merge operations.

Is there a sensorimotor interpretation of Merge consistent with my proposed interpretation of the X-bar schema? My earlier suggestion was that an X-bar schema describes a context-updating deictic operation, activated as part of a deictic routine replayed from working memory: the operation is executed in an initial context, generates a reafferent sensory signal, and results in a new context. If there is a sensorimotor interpretation of Merge, it must reconstrue this replay process, identifying some of its more basic components. I should reiterate that since an LF structure describes a *process* in my interpretation, not a declarative mental representation, the structure formed by Merge will not be interpreted as describing a single static mental representation, constructed from two component mental representations. Rather it will be interpreted as describing a basic unit of spatiotemporal structure in a replayed deictic routine, and its constituents will be interpreted as describing specific sensorimotor signals activated in the course of such a routine. The key question, then is what these signals might be, and what relationships between them might be encoded by the minimal unit of structure created by Merge.

One possible way to interpret the structure formed by a Merge operation is with reference to the associative brain mechanisms that implement the storage of a sequence of deictic operations, and enable its replay. As discussed in Section 7.2, these mechanisms make use of a dynamically updating representation of context. One mechanism associates an initial context representation c with a deictic operation O, so that activating c triggers activation of the operation O. The other associates the operation O—in the current context c—with another sensorimotor signal S. These mechanisms interact: when c becomes active, this activates O; the combination of c and O in turn activates S. Now consider the structure in Figure 10, where a head constituent α is merged with another constituent β. One possible interpretation is that α describes a sensorimotor operation O, β describes some other sensorimotor signal S, and the constituent formed by merging α and β describes the context c, which triggers operation O, and then enables a subsequent association between O and the signal S. This constituent 'represents the combination of α and β' in that the context it describes enables an associative connection between the signals described by α and β. Its being 'labelled' with the head constituent α reflects the fact that the context it describes directly activates the signal described by α. This interpretation makes sense both in the case where the signal S is a reafferent consequence of the deictic operation O, and in the case where S is an updated context representation, which can in turn activate the next deictic operation.

It is beyond the scope of this paper to propose a detailed sensorimotor interpretation of Merge. My main suggestion is that thinking of the structure created by Merge as describing an element of structure in a cognitive *process*, rather than describing a static cognitive representation with component parts, may be helpful in characterising Merge in neural terms. The above interpretation of Merge is not completely satisfactory: it does not quite gel with the proposal that an X-bar schema is derived from two successive applications of Merge. (This would imply that an X-bar schema describes two successive deictic operations, while in the sensorimotor model I am assuming, the deictic operation that activates a reafferent signal is the *same* operation that triggers an updated context representation.) But for the moment, pursuing an improved interpretation will be left as a matter for further work.

10.2 Other Functional Projections in Minimalist Analyses

As also noted in Section 3.2, the LF structure of a clause contains several projections that do not appear in my simple model of LF structure: these include CP, which has a role in the syntax of questions, relative clauses and clausal complements, TP, whose head holds the semantic features signalled by the tense inflections of verbs or tense markers, and more recently vP, a projection headed by a light verb, that introduces the VP proper. If these projections do indeed feature in LF structure—and there is good evidence they do—my general sensorimotor interpretation of right-branching LF structures makes clear predictions that the process of apprehending a episode has additional stages to it, that appear at well-defined serial positions in relation to the stages I have described. I will briefly discuss some ongoing work exploring these predictions.

CP and TP are in the 'left periphery' of LF, above AgrSP; CP is higher than

TP. The existence of these projections predicts that experiencing an episode involves executing two sensorimotor operations in sequence *before attending to the agent*. My basic proposal is that there are several cognitive operations that must be executed in order to put the brain into a state where it is ready to evoke and rehearse a sensorimotor sequence. I will consider the operation corresponding to TP first. My suggestion is that the head of TP describes an operation that determines whether an episode representation is to be retrieved from memory or gathered directly from experience, through the sensorimotor system. As already noted in Section 6.1, there is good evidence that the brain can exist in several distinct modes of connectivity, implemented by large-scale brain networks (Bressler and Menon, 2010). There is a well-studied distinction between a mode associated with memory retrieval and a mode associated with attention to external stimuli (see e.g. Sestieri *et al.*, 2011). While the cognitive operations that establish these alternative modes are not yet well understood, they are good candidates for the operations described by TP, which encodes the distinction between present and past sentences. Turning to CP, I suggest that this projection may describe an even earlier mode-setting cognitive operation. My focus has been on CP as it appears in sentences with sentential complements, such as *X says [that] P*. My proposal is that the complementiser *that* describes a cognitive operation establishing a special 'verbal mode', in which concepts are linked to words rather than to the world. A model of this operation and how infants learn to engage it is given in Caza and Knott (2012); see also Knott (in press). In Knott (2012) I also briefly suggest an interpretation of CP as it appears in questions. In this context, I suggest CP describes an operation that engages the cognitive mode in which queries can be posed to episodic memory. (Again there is good evidence that such a mode exists.) Question formation is traditionally seen as involving the raising of an inflectional head to the head of CP; I interpret this type of head-raising as a reflection of the actual mechanisms through which a query is posed to episodic memory.

I conclude by considering vP. The proposal that VP is introduced by a vP projection headed by a light verb makes a prediction about the process of action monitoring that takes place at the end of a sensorimotor routine. It predicts that action monitoring is more complex than is posited in Section 6.3: rather than being a single continuous process, it should have two well-defined temporal stages that occur in succession. To explore this prediction I have focussed on causative light verbs, originally hypothesised as part of an account of the causative alternation. Verbs undergoing this alternation can appear as transitives but also as intransitives: an example is *John opened the door / The door opened*. A common account of this alternation posits that the LF of *John opened the door* involves two VPs: a higher vP headed by the verb *cause*, introducing a complement VP headed by *open* (thus *John caused [the door opened]*). There is very good evidence for causative actions in the motor system, i.e. for actions that are represented by the perceptual effects they bring about (see Hommel *et al.*, 2001 for a review). Lee-Hand and Knott (2013) present a neural network model of the learning and control of actions defined by their perceptual effects; in this model, executing such an action involves activation of a network that controls a causative action, *and then* monitoring of the perceptual consequences of this action. This sequence of processing corresponds very well to the sequence predicted by the dual-VPs analysis. What is more, the model also offers an interesting sensorimotor interpretation of the special type of head raising that allows the lower verb *open* to raise to adjoin to the light verb *cause*.

There are several other projections posited within LF that remain to be considered: VoiceP, AspP and many others. Testing predictions about these projections is a matter for further work.

I will conclude by revisiting the agreement projections AgrSP and AgrOP. Several theorists have suggested that these can be dispensed with given the presence of other functional projections in the clause: for instance, it has been proposed that the head of vP can check accusative Case and the verb's object agreement features, and that TP can check the verb's subject agreement features (see e.g. the discussion in Hornstein *et al.*, 2005:162–8). Given we do not yet have a well worked out sensorimotor interpretation of vP and TP it is premature to assess these proposals in any detail—but if the above proposals about vP and TP are on the right track, then I suggest it is unlikely the sensorimotor interpretations of these projections overlap with those of AgrSP and AgrOP. For instance, in Lee-Hand and Knott's model of causative actions, the sensorimotor routine involved in executing the action of breaking a cup involves an action of attention to the cup (corresponding to AgrOP) and *then* activation of the causative action network (corresponding to vP). Discrepancies of this kind can push in two directions. On one hand they can indicate problems for the proposed sensorimotor interpretation of LF. On the other hand they can prompt further efforts to motivate a separate AgrOP projection through syntactic argumentation. At present it is not clear which direction will predominate.

10.3 A Reductionist Model of Syntax

The linking hypothesis I propose has a reductionist flavour: I want to explain (some) syntactic structures in language as manifestations of nonlinguistic cognitive phenomena. This is a direction which many linguists have pursued, particularly within the field of cognitive linguistics (see e.g. Lakoff, 1987; Langacker, 1987, 2008). But there are also phenomena which appear to be irreducibly syntactic. A good discussion about the limits of reductionist accounts of syntax is given by Jackendoff (2002). To take a small example, verbs with apparently similar semantics can introduce different prepositions (e.g. *count <u>on</u>* vs *trust <u>in</u>*). It is hard to see such differences as reflecting semantic differences. To take a more substantial example, transitive verbs are able to express a wide range of semantic structures. My cup-grabbing sentence features a verb which takes an agent and a patient, but other transitive verbs take apparently different argument types: *like* takes an experiencer and a stimulus, *frighten* takes a stimulus and an experiencer, *own* takes an 'owner' and an 'ownee' and so on. These verbs all project the same syntactic structure as 'grab', featuring positions for a subject and an object. But do they share the same semantic structure?

To begin with I should note that my model of 'surface language' certainly allows for surface syntactic idiosyncracies which have no origin in semantics, and allows that these idiosyncracies play a large part in the grammar of a language. The sentence generation network described in Section 9 can certainly learn the kinds of arbitrary dependency which feature in phrasal verbs like *count on* and *trust in*.

The question about apparent semantic diversity in the structures projected by *like*, *frighten*, *own* etc is more telling. If these verbs are syntactically identical, and syntactic structures are understood to reflect semantic structures, what semantic characterisation of subject and object position can we give which is general enough to apply to all these semantically disparate verbs?

My main response here is that interpreting a syntactic structure as describing stages in a sensorimotor routine is not the same as seeing it as reflecting a specific semantic pattern. I would certainly agree that different transitive verbs describe eventualities of very different types, and even that these are apprehended through operations in different cognitive modalities. All I am proposing is that there are regularities in the temporal structure of these operations, and that these are reflected in the structure of working memory episode representations, and as a result, in syntax.

The idea that subjects and objects are best defined with reference to the perceptual processing of an episode rather than to its intrinsic semantic structure has been suggested several times before. A common idea is that subjects describe participants with higher 'attentional prominence'; see for instance Langacker (2008); Talmy (2000); also Dowty (1991). One difference in my proposal is that perceptual processes are seen as having strong sequential structure, and subject and object are defined in relation to this structure rather than simply in relation to prominence. (In my account, both agent and patient become prominent, but at different times.) Another difference in my proposal is that it characterises subject and object positions at a language-independent level of syntactic representation (LF) rather than in surface sentence structure. This is helpful in accounting for the argument patterns of stimulus-experiencer verbs like *frighten*, which pose problems for most attempts to characterise argument positions semantically. If our semantic characterisations of subject and object are about positions at LF rather than PF, then we can account for such cases by arguing that the surface object appears higher than the surface subject *at LF*—and there are some good arguments along these lines (see e.g. Pesetsky, 1995; Anagnostopoulou, 1999). Of course these arguments rescue a theory of argument linking at the expense of more complex hypotheses about LF structures—but at least in my approach these hypotheses make predictions about sensorimotor processing which can be independently tested.

10.4 The Idea of Language-Independent LF Structures

My sensorimotor interpretation of LF structures also appears to make a very strong claim about the language-independence of LF. Sensorimotor processes are uncontroversially language-independent, but no Minimalist would want to claim that LF structures are fully invariant over translation: to take a famous example, *John swam across the river* must translate in French to *John traversa le fleuve en nageant* ('John crossed the river swimming'), which has a clearly different LF structure. More relevant to my cup-grabbing example, there are languages where transitive motor actions are easily or even canonically expressed in passive constructions (languages with ergative characteristics like Māori sometimes have this character; see e.g. Harlow, 2007). Where does this leave my sensorimotor account of LF?

Of course even within a single language there are often several alternative ways of expressing an episode syntactically. Any model of language semantics must rely heavily on an inference mechanism, which identifies commonalities in the semantic contributions of sentences which paraphrase one another, such as active and passive versions of a sentence. In relation to the sensorimotor model, this inference mechanism can perhaps be identified with the mechanism which updates an agent's representation of the current context when a deictic routine is completed. I assume an agent's representation of context is a rich, high-dimensional structure, which cannot be directly expressed in language. (In my model, some of the signals which provide input to this update operation have direct linguistic reflexes, but the update operation itself is complex, and learned through long sensorimotor experience.) If we allow that several sensorimotor routines can bring about roughly the same context update operation, then perhaps we can account for cases where LF structures are not preserved across languages by positing that languages can encode conventions about the routines through which particular updates are communicated. It is likely that a sizeable portion of the grammar of any language would have to be made up of conventions of this kind—and this portion of the grammar will probably have an 'empirist', construction-based flavour. But note that the conventions encoded in any given language will not be entirely arbitrary; they will have their origins in the alternative sensorimotor routines through which a given episode can be experienced, and which result in a given update. So a study of the sensorimotor system is still of use in identifying the alternative constructions from which conventions can be formed.

10.5 Deictic Routines

Another foundational assumption of my proposal is that sensorimotor processes are all structured as deictic routines. Does this idea stand up to scrutiny? Are sensorimotor processes structured as sequences at a certain timescale?

Of course in many ways there is massive parallelism in sensorimotor mechanisms. The account of deictic routines which I propose is quite consistent with this. In my model, the deictic operations involved in experiencing an episode progressively extend a neural circuit in which there is continuous and parallel processing. For instance, in the cup-grabbing example, when the observer attends to the agent, he initiates processing in a neural circuit which tracks the agent; when he attends to the cup, he initiates processing in a second circuit which tracks the cup, which is active in parallel with the first circuit, and in which makes use of the representations it generates. When the observer monitors the grab action, this initiates processing in a third circuit, which uses the representations generated by the first two circuits and runs in parallel to them. I also assume that there is parallel activity in neural circuits before they are selected. For instance, when the observer is deciding whether to engage the action-perception circuit or the action-execution circuit, we expect there to be activity in both these circuits concurrently, representing their claims to be selected.

The model of deictic routines nonetheless proposes that there are discrete changes in the neural circuitry active during the apprehension of concrete events. In Ballard *et al.*'s original conception, this general idea was mainly supported by analyses of discrete elements of behaviour, in particular saccadic eye movements.

But it is also supported by the recent discovery of large-scale brain networks (Bressler and Menon, 2010), which are activated or deactivated as wholes by distinct neural operations (Sridharan *et al.*, 2008; Menon and Uddin, 2010), and which appear to have hierarchical structure (Doucet *et al.*, 2011).

10.6 Beyond Concrete Sentences

I have proposed a hypothesis linking the LF structure of concrete sentences to sensorimotor mechanisms. But there are abstract sentences with the very same LF as my example cup-grabbing sentence, for instance *The company acquired a subsidiary*: in this case, the LF structure cannot describe a sensorimotor process in any direct way. Clearly my hypothesis about concrete sentences commits me to some related claim about sentences like these.

A well-known approach taken by linguists interested in embodiment is to argue that abstract sentences aquire their meaning through metaphors grounded in concrete domains (see classically Lakoff and Johnson, 1980). Thus for instance if 'company' is metaphorically an agent and 'subsidiary' is metaphorically an object, then 'acquire' can have a meaning similar to the motor action 'take' or 'get'. If this idea is correct, then the sensorimotor interpretation of LF could possibly be extended to abstract sentences by proposing that the propositions they describe are apprehended and stored as deictic routines, operating over conceptualised or simulated objects rather than actual objects in the world. However, my feeling is that formulating a detailed hypothesis about how abstract propositions are grounded metaphorically in concrete domains first requires a thorough understanding of how concrete episodes are apprehended—which we are far from attaining. As a point of methodology, therefore, I think it may be premature to attempt a detailed metaphor-based account of abstract sentences.

Another way of looking beyond concrete sentences is to consider sentences that do more than report experienced episodes. Sentences can express desires, ask questions, report memories, and do many other things: these capabilities can be traced to particular XPs at LF, and ultimately my interpretation of LF should extend to these XPs too. In order to move in this direction, a natural strategy is to broaden the concept of a 'deictic operation', which currently applies only to sensory and motor actions, so that it includes other types of cognitive operation—for instance, operations which manipulate working-memory representations or which perform storage or retrieval of material in long-term memory. Our interpretation of the X-bar schema would then allow XPs to describe cognitive operations of this kind as well as sensorimotor operations. If there are XPs which can plausibly be interpreted as signalling purely cognitive operations, the LF structures in which they appear may provide interesting ways of thinking about the sequential organisation of these operations. The proposed interpretations of CP and TP discussed in Section 10.2 in fact move in this direction.

11 Summary

In this paper I have made a suggestion about how the rich and complex information received by the senses during the apprehension of a simple reach-to-grasp action is compressed into a linguistic representation. Obviously there is a huge amount of compression; my main suggestion is that the basis for this compression is the temporal structure of the sensorimotor processes—specifically, its structure as a deictic routine. Whether this idea can be successfully extended beyond the example cup-grabbing scenario is a matter for further work.

The proposal I have made about the interface between language and the sensorimotor is expressed in terms of Chomsky's Minimalist model. But the proposal requires some quite radical reinterpretations to Minimalism, particularly of its account of the derivation of LF structures. These revisions allow the Minimalist model to be supplemented with accounts of sentence processing, surface syntax and syntactic development derived from empiricist models of language which are normally regarded as Minimalism's competitors. I still maintain that the Mimimalist conception of LF provides a very helpful framework for a strong hypothesis about how syntactic structures relate to neural mechanisms. But at the same time, this hypothesis may help to restate some of the key insights of Minimalism in a way which is more compatible with alternative conceptions of syntax which are currently more widespread within cognitive science.

References

Abney, S. 1987. *The English Noun Phrase in its Sentential Aspect*. Cambridge, MA: MIT dissertation.

Abraham, W, B Logan, J Greenwood & M Dragunow. 2002. Induction and experience-dependent consolidation of stable long-term potentiation lasting months in the hippocampus. *Journal of Neuroscience* 22. 9626–9634.

Alexiadou, A, L Haegeman & M Stavrou. 2007. *Noun Phrase in the Generative Perspective*. Berlin: Mouton de Gruyter.

Anagnostopoulou, E. 1999. On experiencers. In A Alexiadou, G Horrocks & M Stavrou (eds.), *Studies in Greek Syntax*, 67–93. Dordrecht: Kluwer.

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

Averbeck, B & D Lee. 2007. Prefrontal correlates of memory for sequences. *Journal of Neuroscience* 27(9). 2204–2211.

Baddeley, A. 2000. The episodic buffer: A new component of working memory? *TICS* 4(11). 417–423.

Baddeley, A & G Hitch. 1974. Working memory. In G Bower (ed.), *The psychology of learning and motivation*, 48–79. Academic Press.

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

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

Barsalou, L. 2008. Grounded cognition. *Annual Review of Psychology* 59. 617–645.

Barsalou, L, W Simmons, A Barbey & C Wilson. 2003. Grounding conceptual knowledge in modality-specific systems. *Trends in Cognitive Sciences* 7(2). 84–91.

Beiser, D & J Houk. 1998. Model of cortical-basal ganglionic processing: Encoding the serial order of sensory events. *Journal of Neurophysiology* 79(6). 3168–3188.

Bickerton, D. 1995. *Language and Human Behavior*. London: University College London Press.

Bookheimer, S. 2002. Functional MRI of language: New approaches to understanding the cortical organization of semantic processing. *Annual Review of Neuroscience* 25. 151–188.

Bressler, S & V Menon. 2010. Large-scale brain networks in cognition: emerging methods and principles. *Trends in Cognitive Sciences* 14. 277–290.

Caza, G & A Knott. 2012. Pragmatic bootstrapping: A neural network model of vocabulary acquisition. *Language Learning and Development* 8. 1–23.

Chambon, V, D Wenke, S Fleming, W Prinz & P Haggard. 2012. An online neural substrate for sense of agency. *Cerebral Cortex* 23. 1031–1037.

Chang, F. 2002. Symbolically speaking: A connectionist model of sentence production. *Cognitive Science* 26. 609–651.

Chomsky, N. 1965. *Aspects of the Theory of Syntax*. Cambridge, MA: MIT Press.

Chomsky, N. 1970. Remarks on nominalization. In R Jacobs & P Rosenbaum (eds.), *Reading in English Transformational Grammar*, 184–221. Waltham: Ginn.

Chomsky, N. 1981. *Lectures on Government and Binding*. Dordrecht: Foris.

Chomsky, N. 1995a. Bare phrase structure. In H Campos & P Kempchinsky (eds.), *Evolution and Revolution in Linguistic Theory. Essays in honor of Carlos Otero*, Washington, DC: Georgetown University Press.

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

Christiansen, M & N Chater. 1999. Toward a connectionist model of recursion in human linguistic performance. *Cognitive Science* 23. 157–205.

Cisek, P & J Kalaska. 2010. Neural mechanisms for interacting with a world full of action choices. *Annual Review of Neuroscience* 33. 269–298.

Cohen, J, E Berkman & M Lieberman. 2013. Intentional and incidental self-control in ventrolateral PFC. In D Stuss & R Knight (eds.), *Principles of Frontal Lobe Functions: Second Edition*, 417–440. Cambridge, MA: MIT Press.

Colby, C & M Goldberg. 1999. Space and attention in parietal cortex. *Annual Review of Neuroscience* 22. 319–349.

Corbett, G. 1991. *Gender*. Cambridge/New York: Cambridge University Press.

Critchley, H, C Mathias & R Dolan. 2003. Neuroanatomical basis for first- and second-order representations of bodily states. *Nature Neuroscience* 4(2). 207–212.

Damasio, A (ed.). 1999. *The Feeling of What Happens: Body and Emotion in the Making of Consciousness*. New York: Harcourt Brace.

di Pellegrino, G, L Fadiga, L Fogassi, V Gallese & G Rizzolatti. 1992. Understanding motor events: A neurophysiological study. *Experimental Brain Research* 91. 176–180.

Diba, K & G Buzsàki. 2007. Forward and reverse hippocampal place-cell sequences during ripples. *Nature Neuroscience* 10(10). 1241–1242.

Domahs, F, A Nagels, U Domahs, C Whitney, R Wiese & T Kircher. 2012. Where the mass counts: Common cortical activation for different kinds of nonsingularity. *Journal of Cognitive Neuroscience* 24(4). 915–932.

Dominey, P, M Arbib & J-P Joseph. 1995. A model of corticostriatal plasticity for learning associations and sequences. *Journal of cognitive neuroscience* 7(3). 311–336.

Doucet, G, M Naveau, L Petit, N Delcroix, L Zago, F Crivello, G Jobard, N Tzourio-Mazoyer, B Mazoyer, E Mellet & M Joliot. 2011. Brain activity at rest: a multiscale hierarchical functional organization. *Journal of Neurophysiology* 105. 2753–2763.

Dowty, D. 1991. Thematic proto-roles and argument selection. *Language* 67(3). 547–619.

Eichenbaum, H. 2004. Hippocampus: cognitive processes and neural representations that underlie declarative memory. *Neuron* 44. 109–120.

Elman, J. 1990. Finding structure in time. *Cognitive Science* 14. 179–211.

Fadiga, L, L Craighero & E Olivier. 2005. Human motor cortex excitability during the perception of others' action. *Current opinion in neurobiology* 15(2). 213–218.

Fagg, A & M Arbib. 1998. Modeling parietal-premotor interactions in primate control of grasping. *Neural Networks* 11(7/8). 1277–1303.

Farrer, C, S Frey, J Van Horn, E Tunik, D Turk, S Inati & S Grafton. 2008. The angular gyrus computes action awareness representations. *Cerebral Cortex* 18. 254–261.

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

Ferri, F, G Campione, R Dalla Volta, C Gianelli & M Gentilucci. 2010. To me or to you? when the self is advantaged. *Experimental Brain Research* 203. 637–646.

Fillmore, C, P Kay & M O'Connor. 1988. Regularity and idiomaticity in grammatical constructions: The case of *let alone*. *Language* 64. 501–38.

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

Foster, D & M Wilson. 2006. Reverse replay of behavioural sequences in hippocampal place cells during the awake state. *Nature* 440. 680–683.

Freedman, D, M Riesenhuber, T Poggio & E Miller. 2003. A comparison of primate prefrontal and inferior temporal cortices during visual categorisation. *Journal of Neuroscience* 15. 5235–5246.

Gallese, V, L Fadiga, L Fogassi & G Rizzolatti. 1996. Action recognition in the premotor cortex. *Brain* 119. 593–609.

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

Geyer, S, M Matelli, G Luppino & K Zilles. 2000. Functional neuroanatomy of the primate isocortical motor system. *Anatomy and Embryology* 202(6). 443–474.

Glenberg, A & P Kaschak. 2002. Grounding language in action. *Psychonomic Bulletin and Review* 9(3). 558–565.

Goldberg, A. 1995. *Constructions. A Construction Grammar approach to Argument Structure*. Chicago: University of Chicago Press.

Goldstein, L & C Fowler. 2003. Articulatory phonology: A phonology for public language use. In A Meyer & N Schiller (eds.), *Phonetics and Phonology in Language Comprehension and Production: Differences and Similarities*, 159–207. New York: Mouton.

Goodwin, A & H Wheat. 2004. Sensory signals in neural populations underlying tactile perception and manipulation. *Annual Review of Neuroscience* 27. 53–77.

Grèzes, J & J Decety. 2001. Functional anatomy of execution, mental simulation, observation and verb generation of actions: A meta-analysis. *Human Brain Mapping* 12(1). 1–19.

Haggard, P. 2008. Human volition: Towards a neuroscience of will. *Nature Reviews Neuroscience* 9. 934–946.

Hamker, F. 2004. A dynamic model of how feature cues guide spatial attention. *Vision Research* 44. 501–521.

Harlow, R. 2007. *Māori: A Linguistic Introduction*. Cambridge, UK: Cambridge University Press.

Hasegawa, R, M Matsumoto & A Mikami. 2000. Search target selection in monkey prefrontal cortex. *Journal of Neurophysiology* 84. 1692–1696.

Henson, R, N Burgess & C Frith. 2000. Recoding, storage, rehearsal and grouping in verbal short-term memory: An fMRI study. *Neuropsychologia* 38(4). 426–440.

Hommel, B, J Müsseler, G Aschersleben & W Prinz. 2001. The theory of event coding (TEC): A framework for perception and action learning. *Behavioral and Brain Sciences* 24. 849–878.

Hornstein, N. 2009. *A Theory of Syntax: Minimal Operations and Universal Grammar*. Cambridge, UK: Cambridge University Press.

Hornstein, N, J Nunes & K Grohmann. 2005. *Understanding Minimalism*. Cambridge, UK: Cambridge University Press.

Houghton, G & T Hartley. 1995. Parallel models of serial behaviour: Lashley revisited. *Psyche* 2(25).

Jackendoff, R. 1977. *X-Bar syntax: A study of phrase structure.* Linguistic Inquiry Monograph 2. Cambridge, MA: MIT Press.

Jackendoff, R. 1999. Possible stages in the evolution of the language capacity. *Trends in Cognitive Sciences* 3(7). 272–279.

Jackendoff, R. 2002. *Foundations of Language: Brain, Meaning, Grammar, Evolution*. Oxford: Oxford University Press.

Jeannerod, M. 1996. Reaching and grasping. Parallel specification of visuomotor channels. In W Prinz & B Bridgeman (eds.), *Handbook of Perception and Action Volume II: Motor Skills*, 405–460. Academic Press.

Jeannerod, M. 2001. Neural simulation of action: A unifying mechanism for motor cognition. *NeuroImage* 14. S103–S109.

Jellema, T, C Baker, B Wicker & D Perrett. 2000. Neural representation for the perception of the intentionality of actions. *Brain and Cognition* 44. 280–302.

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

Kayne, R. 1994. *The Antisymmetry of Syntax*. Cambridge, MA: MIT Press.

Knott, A. 2012. *Sensorimotor Cognition and Natural Language Syntax*. Cambridge, MA: MIT Press.

Knott, A. in press. How infants learn word meanings and propositional attitudes: a neural network model. In T-W Hung (ed.), *Language and Action*,

Berlin/Heidelberg: Springer.
Lakoff, G. 1987. *Women, Fire and Dangerous Things*. Chicago and London: University of Chicago Press.
Lakoff, G & M Johnson. 1980. *Metaphors We Live By*. Chicago and London: University of Chicago Press.
Land, M & S Furneaux. 1997. The knowledge base of the oculomotor system. *Philosophical Transactions of the Royal Society of London B* 352. 1231–1239.
Langacker, R. 1987. Nouns and Verbs. *Language* 63(1). 53–94.
Langacker, R. 2008. *Cognitive Grammar: A Basic Introduction*. New York: Oxford University Press.
Lee-Hand, J & A Knott. 2013. A model of causative actions: from motor learning to syntactic structure. In *Proceedings of the 35th Annual Meeting of the Cognitive Science Society*, Berlin.
Levelt, W. 1989. *Speaking: From Intention to Articulation*. Cambridge, MA: MIT Press.
Levelt, W, A Roelofs & A Meyer. 1999. A theory of lexical access in speech production. *Behavioural and Brain Sciences* 22(1). 1–75.
Lieven, E, J Pine & G Baldwin. 1997. Lexically-based learning and early grammatical development. *Journal of Child Language* 24. 187–219.
MacWhinney, B. 2005. Item-based constructions and the logical problem. In *Proceedings of the Workshop on Psychocomputational Models of Human Language Acquisition*, 53–68.
Marantz, A. 2005. Generative linguistics within the cognitive neuroscience of language. *Linguistic Review* 22. 429–445.
McClelland, J, B McNaughton & R O'Reilly. 1995. Why there are complementary learning systems in the hippocampus and neocortex: Insights from the successes and failures of connectionist models of learning and memory. *Psychological Review* 102. 419–457.
Menon, V & L Uddin. 2010. Saliency, switching, attention and control: a network model of insula function. *Brain Structure and Function* 214. 655–667.
Meteyard, L, S Rodriguez Cuadrado, B Bahrami & G Vigliocco. 2012. Coming of age: A review of embodiment and the neuroscience of semantics. *Cortex* 48. 788–804.
Murata, A, V Gallese, G Luppino, M Kaseda & H Sakata. 2000. Selectivity for the shape, size, and orientation of objects for grasping in neurons of monkey parietal area AIP. *Journal of Neurophysiology* 83(5). 2580–2601.
Novick, J, J Trueswell & S Thompson-Schill. 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 & M Arbib. 2002. Schema design and implementation of the grasp-related mirror neuron system. *Biological Cybernetics* 87. 116–140.
Oztop, E, N Bradley & M Arbib. 2004. Infant grasp learning: a computational model. *Experimental Brain Research* 158. 480–503.
Oztop, E, D Wolpert & M Kawato. 2005. Mental state inference using visual control parameters. *Cognitive Brain Research* 22. 129–151.
Pantel, P & D Lin. 2002. Discovering word senses from text. In *Proceedings of the ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, 613–619.
Perani, D, S Cappa, T Schnur, M Tettamanti, S Collina, M Rosa & F Fazio. 1999.

The neural correlates of verb and noun processing—A PET study. *Brain* 122. 2337–2344.

Pesetsky, D. 1995. *Zero Syntax: Experiencers and Cascades*. Cambridge: MIT Press.

Petreska, B & A Billard. 2009. Movement curvature planning through force field internal models. *Biological Cybernetics* 100. 331–350.

Plate, T. 2003. *Holographic Reduced Representations* CSLI Lecture Notes Number 150. Stanford, CA: CSLI Publications.

Poeppel, D & D Embick. 2005. The relation between linguistics and neuroscience. In A Cutler (ed.), *Twenty-First Century Psycholinguistics: Four Cornerstones*, Hillsdale, NJ: Lawrence Erlbaum.

Pollard, C & I Sag. 1994. *Head-Driven Phrase Structure Grammar*. University of Chicago Press.

Pulvermüller, F, W Lutzenberger & H Preissl. 1999. Nouns and verbs in the intact brain: Evidence from event-related potentials and high-frequency cortical responses. *Cerebral Cortex* 9(5). 497–506.

Pulvermüller, F. 2010. Brain embodiment of syntax and grammar: Discrete combinatorial mechanisms spelt out in neuronal circuits. *Brain and Language* 112. 167–179.

Ramsey, S & A Hamilton. 2010. Understanding actors and object-goals in the human brain. *NeuroImage* 50. 1142–1147.

Rhodes, B, D Bullock, W Verwey, B Averbeck & M Page. 2004. Learning and production of movement sequences: Behavioral, neurophysiological, and modeling perspectives. *Human Movement Science* 23. 699–746.

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

Sahin, N, S Pinker, S Cash, D Schomer & E Halgren. 2009. Sequential processing of lexical, grammatical, and phonological information within Broca's area. *Science* 326. 445–449.

Sestieri, C, M Corbetta, G Romani & G Shulman. 2011. Episodic memory retrieval, parietal cortex, and the default mode network: Functional and topographic analyses. *Journal of Neuroscience* 31(12). 4407–4420.

Shapiro, K & A Caramazza. 2003. The representation of grammatical categories in the brain. *Trends in Cognitive Sciences* 7(5). 201–206.

Shapiro, K, L Moo & A Caramazza. 2012. Neural specificity for grammatical operations is revealed by content-independent fMR adaptation. *Frontiers in Psychology* 3. Doi: 10.3389/fpsyg.2012.00026.

Shastri, L. 2001. Episodic memory trace formation in the hippocampal system: A model of cortico-hippocampal interaction. Tech. Rep. TR-01-004 International Computer Science Institute (ICSI), UC Berkeley Berkeley, CA.

Shastri, L. 2002. Episodic memory and cortico-hippocampal interactions. *Trends in Cognitive Sciences* 6(4). 162–168.

Shima, K & J Tanji. 2000. Neuronal activity in the supplementary and presupplementary motor areas for temporal organization of multiple movements. *Journal of Neurophysiology* 84(4). 2148–2160.

Sridharan, D, D Levitin & V Menon. 2008. A critical role for the right fronto-insular cortex in switching between central-executive and default-mode networks. *Proceedings of the National Academy of Sciences of the United States of America*

105(34). 12569–12574.
Steedman, M. 2002. Plans, affordances and categorial grammar. *Linguistics and Philosophy* 25. 723–753.
Swallow, K, J Zacks & R Abrams. 2009. Event boundaries in perception affect memory encoding and updating. *Journal of Experimental Psychology: General* 138(2). 236–257.
Takac, M, L Benuskova & A Knott. 2012. Mapping sensorimotor sequences to word sequences: A connectionist model of language acquisition and sentence generation. *Cognition* 125. 288–308.
Takac, M & Knott. 2013. A neural network model of working memory for episodes. In *Proceedings of the 35th Annual Meeting of the Cognitive Science Society*, Berlin.
Talmy, L. 2000. *Towards a Cognitive Semantics. Volume 1: Concept Structuring Systems*. MIT Press.
Tanenhaus, M, M Spivey-Knowlton, K Eberhard & J Sedivy. 1995. Integration of visual and linguistic information in spoken language comprehension. *Science* 268. 1632–1634.
Tomasello, M. 2003. *Constructing a Language: A Usage-Based Theory of Language Acquisition*. Harvard University Press.
Tomita, H, M Ohbayashi, K Nakahara & Y Hasegawa, I Miyashita. 1999. Top-down signal from prefrontal cortex in executive control of memory retrieval. *Nature* 401. 699–703.
Tranel, D, R Adolphs, H Damasio & A Damasio. 2001. A neural basis for the retrieval of words for actions. *Cognitive Neuropsychology* 18(7). 655–674.
van der Velde, F & M de Kamps. 2006. Neural blackboard architectures of combinatorial structures in cognition. *Behavioral and Brain Sciences* 29. 37–108.
Wallenstein, G, H Eichenbaum & M Hasselmo. 1998. The hippocampus as an associator of discontiguous events. *Trends in Neurosciences* 21. 317–323.
Walles, H. 2010. *A Perceptual Basis for Noun Phrase Syntax*: Department of Computer Science, University of Otago dissertation.
Webb, A, A Knott & M MacAskill. 2010. Eye movements during transitive action observation have sequential structure. *Acta Psychologica* 133. 51–56.
Wray, A. 1998. Protolanguage as a holistic system for social interaction. *Language and Communication* 18. 47–67.
Yamaguchi, S, S Yamagata & S Kobayashi. 2000. Cerebral asymmetry of the "top-down" allocation of attention to global and local features. *Journal of Neuroscience* 20(9). RC72.
Zamparelli, R. 2000. *Layers in the Determiner Phrase*. New York: Garland.
Zwaan, R & L Taylor. 2006. Seeing, acting, understanding: Motor resonance in language comprehension. *Journal of Experimental Psychology: General* 135(1). 1–11.

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