

Dept of Computer Science  
University of Otago  
P.O. Box 56, Dunedin  
New Zealand

November 13, 2013

Dear Prof. Grohmann,

I'd like to thank the two reviewers for their comments on my article 'Syntactic structures as descriptions of sensorimotor processes'. These were extremely helpful. I have revised the article in several ways, along the lines suggested by the reviewers.

Reviewer A's comments are all suggestions about how to update the linguistic framework used in the paper, in line with the current bare-phrase-structure version of Minimalism. I agree with the reviewer that it would be helpful to adopt a contemporary version of Minimalism, and, for the most part, I think it is possible to do this. But since the paper is largely a précis of my recent book, which adopts an early version of Minimalism throughout, I don't want to make a wholesale update in the paper. Instead I have extended the paper in several places with indications about how the syntactic model could be restated in more contemporary Minimalist terms.

1. The reviewer asks that the X-bar schema should be shown to be derivable from the more basic operation Merge. I now note this in Section 3.2. I have added a new section (Section 10.1) discussing whether there is a sensorimotor characterisation of Merge that identifies a more primitive combinatorial component of sensorimotor routines. I think this is a really interesting question. I have made one suggestion about a sensorimotor interpretation of Merge—it is not fully satisfactory, but I think it's an interesting basis for further work.
2. The reviewer asks about the many recent additional projections that are now assumed to feature in LF structure (vP, TP, VoiceP, AspP and

so on). I have noted the omission of these projections more clearly on p6, and I have added a section (Section 10.2) discussing these in more detail.

- (a) I'll consider the split VP first. The reviewer suggests there may be a sensorimotor interpretation of VP shells, compatible with my general sensorimotor interpretation of nested complement structure, whereby the higher vP denotes a motor action and the lower VP denotes its end state. I think this is bang on the money! I have just finished a paper proposing a sensorimotor interpretation of VP shells along exactly these lines. This paper focusses on causative light verbs, and gives a sensorimotor interpretation of the VP-shells structures that linguists posit in an account of the causative alternation. People like Levin and Rappaport suggest that the LF of *John broke the cup* contains nested VPs: '[vP cause [VP break cup]]'. With a student I have developed a computational model of the learning and control of causative actions (Lee-Hand and Knott, 2013; Lee-Hand and Knott, under review), in which the sequence of sensorimotor processes involved in executing actions like 'bend', 'open' and 'squash' can be mapped in detail to a nested VP structure of this kind. I put this paper online—see References for the URL. Any comments on the proposed sensorimotor interpretation of VP shells by the reviewers would be welcome, obviously!
- (b) Next I'll consider CP and TP. I certainly agree that these projections need to feature in an account of the LF of a clause. Given that they feature in LF structure, my hypothesis predicts that there are sensorimotor operations that correspond to them, occurring at appropriate points in the sequence of sensorimotor operations associated with cup-grasping. I think there's some evidence that supports these predictions. I assume that CP and TP are in the left periphery of the LF of a clause, with CP appearing higher than TP. Very briefly, my suggestion is that CP and TP describe cognitive operations that put the brain into a state where it is ready to evoke and rehearse a sensorimotor sequence. CP describes an operation that resets the neural medium that stores propositions in working memory, so it's ready to store a new proposition, while TP describes the cognitive operation that selects the cognitive modality through which the *next* proposition to occupy working memory will be accessed.

- To take TP first: there’s good evidence that there is a cognitive operation that configures the brain to receive material retrieved from episodic memory, and an alternative operation that configures the brain to engage the sensorimotor system and receive information from immediate experience. I want to argue that these operations are signalled by the head of TP, with present tense signalling establishment of experience mode and past tense signalling (episodic) memory mode. These mode-setting operations certainly occur before the attentional and motor operations that represent the content of a proposition, so their position in the left periphery of LF fits with my general hypothesis.
- As for CP: the work I’ve done on that has mainly been on the complementiser that introduces propositional attitudes (*X says [that P]*). My suggestion is that *that* signals a cognitive operation that establishes a cognitive mode specific to humans, where concepts are activated by words rather than by experience or memory. I’ve presented this idea in a couple of papers (Caza and Knott, 2012; Knott, in press).

This work in progress gives some indication that there are sensible things to say about at least some of the more recently proposed functional projections. I don’t think there’s room to go into much detail in the current paper, which is already long, but I have included a section that summarises the work (Section 10.2).

3. Finally, the reviewer suggests I should omit Agr projections from the LF structure proposed for *X grabbed Y*. I appreciate that Agr projections have been dispensed with in recent versions of Minimalism, and other projections used instead to host the features that check Case. The main suggestions I know are 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). Since I’m still working on a sensorimotor interpretation of TP and vP, I’m not yet sure whether these projections can stand in for AgrSP and AgrOP—but my feeling is that the sensorimotor interpretations I’m envisaging for TP and vP *don’t* overlap with those I’ve given to AgrSP and AgrOP. For instance, my proposal about the LF of *John broke the cup* is that it contains the structure ‘[AgrOP Agr [vP cause [VP break cup]]]’: in sensorimotor terms, AgrO signals the action of attending to the cup (and its speci-

fier signals the reafferent sensory representation of the cup); vP signals the causative motor action (and its specifier signals the reafferent representation of the agent as a pattern of movement). These really are separate sensorimotor operations, so my prediction is for two separate projections. Of course, these projections have to be justifiable through linguistic argumentation as well. I can see that there are arguments for dispensing with AgrOP. But I'm not sure these are conclusive. For instance, they rely on the possibility of categories having more than one specifier, which is not uncontroversial. And there are also arguments for retaining both TP and AgrSP—for instance the Icelandic expletive construction discussed in Hornstein *et al.* (2005:167-8. In any case, I've included a brief discussion of the status of Agr projections at the end of Section 10.2.

One final point—I agree with the reviewer that I'm using Agr projections somewhat nonstandardly, in that they carry real semantic content. I appreciate that AgrS and AgrO as originally proposed carried uninterpretable features, so their projections were not visible at the LF interface. In my model, Agr features have as much meaning as other heads, in that they denote attentional operations. Actually I think this strengthens the case for Agr projections... but it's nonetheless important to be able to justify these projections purely through syntactic argumentation. I've noted these points in Section 8.3.

Reviewer A also notes that LF has no linear order in the standard version of Minimalism, and therefore that it's inappropriate to talk about LF structures as being right-branching. I agree that the term is inappropriate, strictly speaking, but it's a convenient shorthand for a structure where recursion happens through repeated adjunctions to complement rather than specifier positions. I introduce the term more carefully on p5. For most of the paper I continue to use it, in the light of Kayne's proposal that hierarchical structure at LF determines linear order *at PF*. This decision is now justified explicitly on p6.

The changes made in response to Reviewer B's comments are as follows.

1. The reviewer argues that the constructionist/empiricist model of syntactic development has not been shown to be more plausible than the Minimalist model. I certainly agree that there are huge problems with Elman's empiricist model of syntactic development if it's seen as a complete theory by itself. (And some people really do seem to view it

that way.) But I don't think we need to think of the empiricist and Minimalist models as alternatives to one another. One idea I really want to emphasise in the current paper is that if LF is construed in sensorimotor terms, as a description of a simulated sensorimotor process, then there is a natural way of expressing a model of language processing that is compatible with the Minimalist account of syntactic representation, and, building on this, a natural way of modelling the actual learning mechanism that learns a mapping between LF and PF for a given exposure language. This is all described in Section 9.3. The learning mechanism is a neural network; this network *incorporates* an Elman-style simple recurrent network, but it also contains circuitry that implements a very Chomskyan notion of parameter-setting. I'm a reconciliationist! I think knowledge of language is partly knowledge of a repertoire of surface constructions in language, but also partly knowledge of structures which are common to all languages, and which are best modelled using Minimalism. I've tried to express this in the summary (Section 11) as well as in Section 9.3.

Incidentally, I couldn't track down the Frank and Lassota paper re-analysing Tomasello's data. If you provide a citation I'll have another look.

2. The reviewer queries whether my account of concrete transitive action sentences can be extended to abstract sentences in the way proposed by Lakoff and Johnson. I certainly don't want to give the impression I buy into Lakoff and Johnson's account in any detail; I just want to refer to existing work on metaphor as one possible way of extending a sensorimotor model of syntax into the domain of abstract sentences. Actually, as a matter of methodology I don't think we should think about abstract sentences at all, until we have a really well worked out account of the syntax and semantics of concrete sentences. This is a point I make in my book; I have revised the discussion in Section 10.6 to make a similar point.
3. The reviewer notes that there are many concrete sentences beyond transitive action sentences that I should consider. I agree, but I can't consider them all at once! As I say at the start of Section 10, my proposal about the sensorimotor interpretation of a 'right-branching' LF structure makes predictions about many concrete sentences; these predictions basically define a whole programme of research at the interface between syntax and sensorimotor neuroscience. My real aim is to encourage readers to adopt this general research programme, and

use Minimalist analyses of concrete sentences to generate interesting hypotheses about sensorimotor processing that can be independently tested. *The cat is on the mat* is a great example of something my hypothesis makes predictions about: I'm certainly thinking about these, but they're still in the realm of future work!

4. The reviewer makes a very good point about the complexity of the 'grab' motor action. This action is described by a single XP (the VP), but it has a lot of internal complexity—in particular there are parallel neural pathways for the 'reach' and 'grasp' aspects of the action, as the reviewer points out. It is surprising, *prima facie*, that all this complexity is packed into a single discrete 'stage' of the proposed sensorimotor sequence. However, that is indeed what I suggest. There are three reasons why I construe the monitoring (execution or observation) of the 'grab' action as a single stage. Firstly, while it is temporally extended, it is a continuous process; as I discuss in Section 6.3, it has the character of a dynamical system, in which a representation of motor state is continuously updated, until an attractor state (the consequent state of the action) is reached. In terms of temporal structure, there are no obvious staging points during this dynamical process, so it makes sense to consider the process as a whole. Secondly, while there are certainly motor representations that change dynamically while action monitoring is under way, there are also higher-level representations in the motor system that are tonically active, and do not change. The motor programme itself is one such representation. In my model of motor control (see again Lee-Hand and Knott, 2013; Lee-Hand and Knott, under review), the motor programme is selected at the start of action monitoring and remains tonically active without changing until the action is completed: its activity is *responsible* for the continuous dynamics of the hand/arm, but the motor programme is not *itself* dynamically updated. My suggestion is that language interfaces with the tonically active representation of the motor programme, not with the dynamically changing representation of motor state; from the perspective of the linguistic interface, the action monitoring stage has the same discrete character as the earlier attentional operations. Finally, while it's true that reaching-to-grasp involves activity in separate 'reach' and 'grasp' pathways, activity in these pathways occurs in parallel—and the tonically active motor programme influences dynamic activity in both pathways. My suggestion is that it's the *discrete temporal* structure of the sensorimotor routine that's made available

to the linguistic interface, i.e. that's encoded at LF. I have made these points more explicitly at the end of Section 6.3.

Incidentally, I certainly don't want to claim that *all* motor actions are like reaching-to-grasp in being temporally nondecomposable. For instance, I think motor actions like 'break' and 'open' have two well-defined phases (see my comment 2(a) to reviewer A)—and I also think these phases correspond to distinct syntactic constituents at LF. Reaching-to-grasp may be quite unusual in having a uniform structure; most actions are probably more complex.

5. The reviewer notes that in the recognition of actions, the 'grasp' component plays a particularly important role, as shown in the experiments of Costantini and Sinigaglia. No citation was given, but I certainly know of the experiment by Ambrosini, Costantini and Sinigaglia (2011). This shows that the anticipatory eye movements of observers watching an agent reaching for a target are influenced by the agent's grip aperture. I certainly acknowledge that—but the 'reach' component of an action also has an important effect on an observer's anticipatory eye movements during action observation. For instance, when observers watch an agent reaching for a target in the presence of distractors of the same shape, they make anticipatory eye movements to the target based on other cues, in particular the agent's gaze direction and hand trajectory; this was shown in an experiment done in my lab (Webb *et al.*, 2010, cited in a later paper by Costantini *et al.*, 2012). The reviewer notes that it's important in my model that the sensorimotor routine for action observation should match that for action execution; that's absolutely right, but I think there's good evidence for parallel activity in the 'reach' and 'grasp' pathways in action observation, just as there is in action execution. I haven't noted this point in the revised paper, but it's something that's covered in my book (Knott, 2012).
6. My statements about the relation between LF and the sensorimotor system use the word 'describe' (LF structures 'describe' sensorimotor processes). The reviewer queries my use of this word, on the grounds that the relation I'm asserting is structural rather than semantic. I certainly want to say that the structure of LF is similar to the structure of a replayed sensorimotor routine. But I also want to say something about LF as a semantic representation: I want to say that LF functions as a semantic representation (of a concrete sentence) *because* it can be interpreted as a representation of a sensorimotor routine. I al-

luded to this idea in Section 5, in the paragraph after Proposal 1 (the initial statement of LF as a description of a sensorimotor routine): I noted that this statement can be understood as an expression of a simulationist account of meaning within a Minimalist framework. Having said that, I agree it's quite odd to talk of an LF structure as 'describing' a sensorimotor routine. An LF structure is a theoretical device used by linguists: obviously it doesn't describe anything *to* anyone (it's not a communicative device). My problem is that there's no existing term that captures the relation I want to posit between LF structures and sensorimotor routines. I don't want to say that LF structures 'denote' sensorimotor routines, firstly because 'denote' is normally used to indicate model-theoretic representations of meaning, and I don't have those in mind, and secondly because I'm not interpreting LF constituents declaratively, as representations that combine the denotations of their constituents into new expressions (in the way that happens in traditional formal semantics). I don't want to say that LF structures 'represent' sensorimotor routines, because to the linguists who invented them they just don't: they represent sentence structures. My strong preference is to stick with the word 'describe' in the current paper, and simply make a general reference to a simulationist account of meaning by way of explanation; I think too much terminological meta-analysis would get in the way of the actual interpretations of LF being proposed.

7. Finally, the reviewer suggests I should tone down the ambition of the paper, and restrict myself to making a proposal about transitive action sentences, rather than attempting to provide 'a grand theory of everything linguistic'. I certainly don't want to account for all syntactic structure in sensorimotor terms. As I noted in my response to Reviewer B's first comment, I want to leave room in a Minimalist-inspired model of syntax for the kind of surface-based constructions that feature in empiricist models of language. But I *do* want to make a general claim linking the components of LF to the sensorimotor system. I think that the strength of the Minimalist model is that it sees language as the product of a few very simple, general structure-building principles. I think any attempt to connect these to prelinguistic cognitive processes should be similarly general in scope, or it won't accurately reflect the emphasis on generality that's inherent in Minimalism. So I haven't toned down the general claim. I actually think it's useful to state a general claim, that makes testable predictions

about other concrete sentences.

Again I would like to thank the reviewers for their comments, which I think have substantially improved the paper. I look forward to hearing the reviewers' comments on the revised version!

Yours sincerely

Alistair Knott

\*References

- Ambrosini, E., Costantini, M., and Sinigaglia, C. (2011). Grasping with the eyes. *Journal of Neurophysiology*, **106**, 1437–1442.
- Caza, G. and Knott, A. (2012). Pragmatic bootstrapping: A neural network model of vocabulary acquisition. *Language Learning and Development*, **8**, 1–23.
- Costantini, M., Ambrosini, E., and Sinigaglia, C. (2012). Does how i look at what you're doing depend on what i'm doing? *Acta Psychologica*, **141**, 199–204.
- Hornstein, N., Nunes, J., and Grohmann, K. (2005). *Understanding Minimalism*. Cambridge University Press, Cambridge, UK.
- Knott, A. (2012). *Sensorimotor Cognition and Natural Language Syntax*. MIT Press, Cambridge, MA.
- Knott, A. (in press). How infants learn word meanings and propositional attitudes: a neural network model. In T.-W. Hung, editor, *Language and Action*. Springer, Berlin/Heidelberg. Currently available at <http://www.cs.otago.ac.nz/staffpriv/alik/papers/mentalstates.pdf>.
- Lee-Hand, J. and Knott, A. (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.
- Lee-Hand, J. and Knott, A. (under review). Causative motor actions and causative alternation. Manuscript. Currently available at <http://www.cs.otago.ac.nz/staffpriv/alik/papers/causatives.pdf>.

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