Syntactic structures as traces of sensorimotor event representations

Alistair Knott
Dept of Computer Science
University of Otago

Modelling sentences/sentence-sized events

Consider a sentence describing a simple transitive action:

The man grabbed a cup.

Modelling sentences/sentence-sized events

Consider a sentence describing a simple transitive action:

The man grabbed a cup.

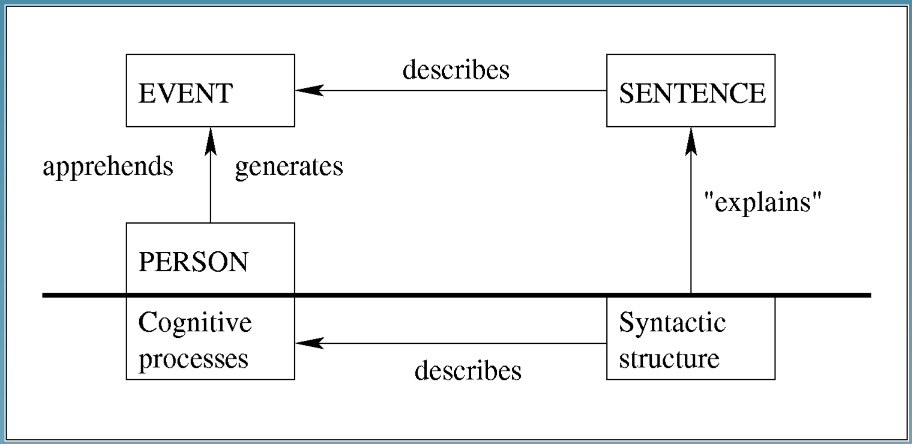
- Linguists try to determine the syntactic and semantic structure of the sentence.
- Psychologists try to model how the action described can be recognised, represented, remembered, executed.

A hypothesis

- A transitive sentence describes an action. (Informally.)
- The syntactic structure of the sentence describes the sensorimotor processes involved in perceiving or executing this action. (More precisely.)

A hypothesis

- A transitive sentence describes an action. (Informally.)
- The syntactic structure of the sentence describes the sensorimotor processes involved in perceiving or executing this action. (More precisely.)



Outline of the talk

1. A model of ‘sentence-sized’ episodes of sensorimotor processing.
2. A model of **memory**, for learning and recalling such episodes.
3. A model of **reinforcement learning**, for generating new such episodes.
4. A model of sentence syntax.
5. The precise version of the hypothesis.

Perceiving sentence-sized events

The process of perceiving our example 'grab' action *decomposes* into several interacting sub-processes.

- Direction of attention to a point in space, and identification of a man at this point.
- Direction of attention to another point in space, and identification of a cup at this point.
- Tracking of the man's movements; identification of the 'grab' motor programme; recognition of its success.

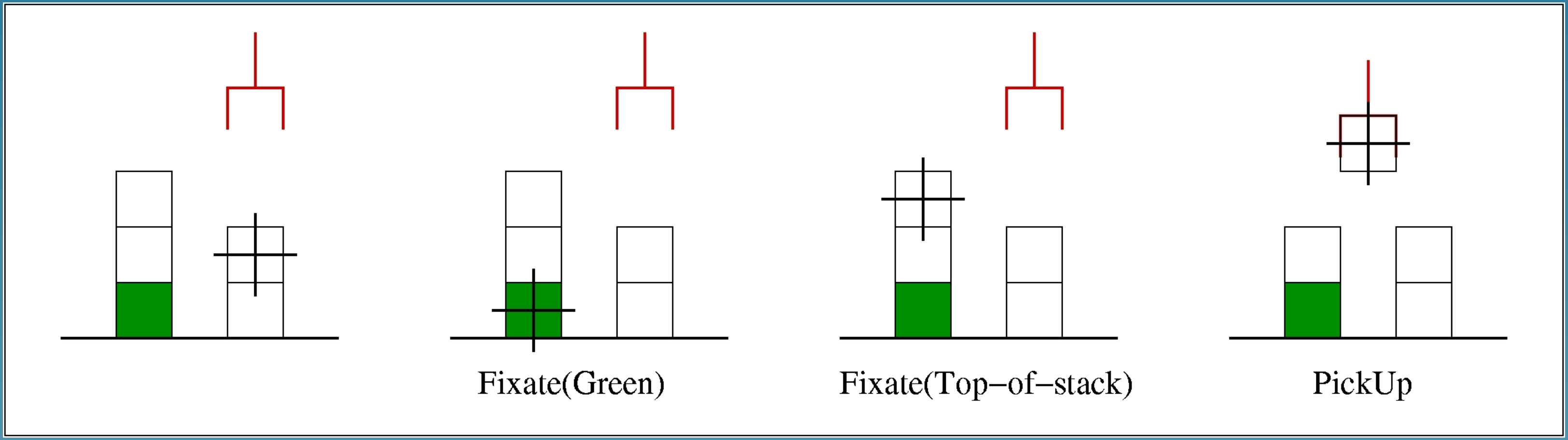
Serial composition of sub-processes

Ballard *et al.*'s (1997) model of **deictic routines**:

- A (human) observer can direct attention to an external point, and create **deictic representations**, which are implicitly referred to this point.
- Deictic representations bind cognitive routines (e.g. directions of attention, motor programmes) to objects in the world.
- Deictic primitives simplify complex behaviours, because each primitive defines the context for its successor.

An example of a deictic routine

(Start of) a sequence for picking up a green block:



The cognitive routines:

Fixate(Colour); Fixate(Relative_Position);
PickUp; PutDown.

Learning deictic routines

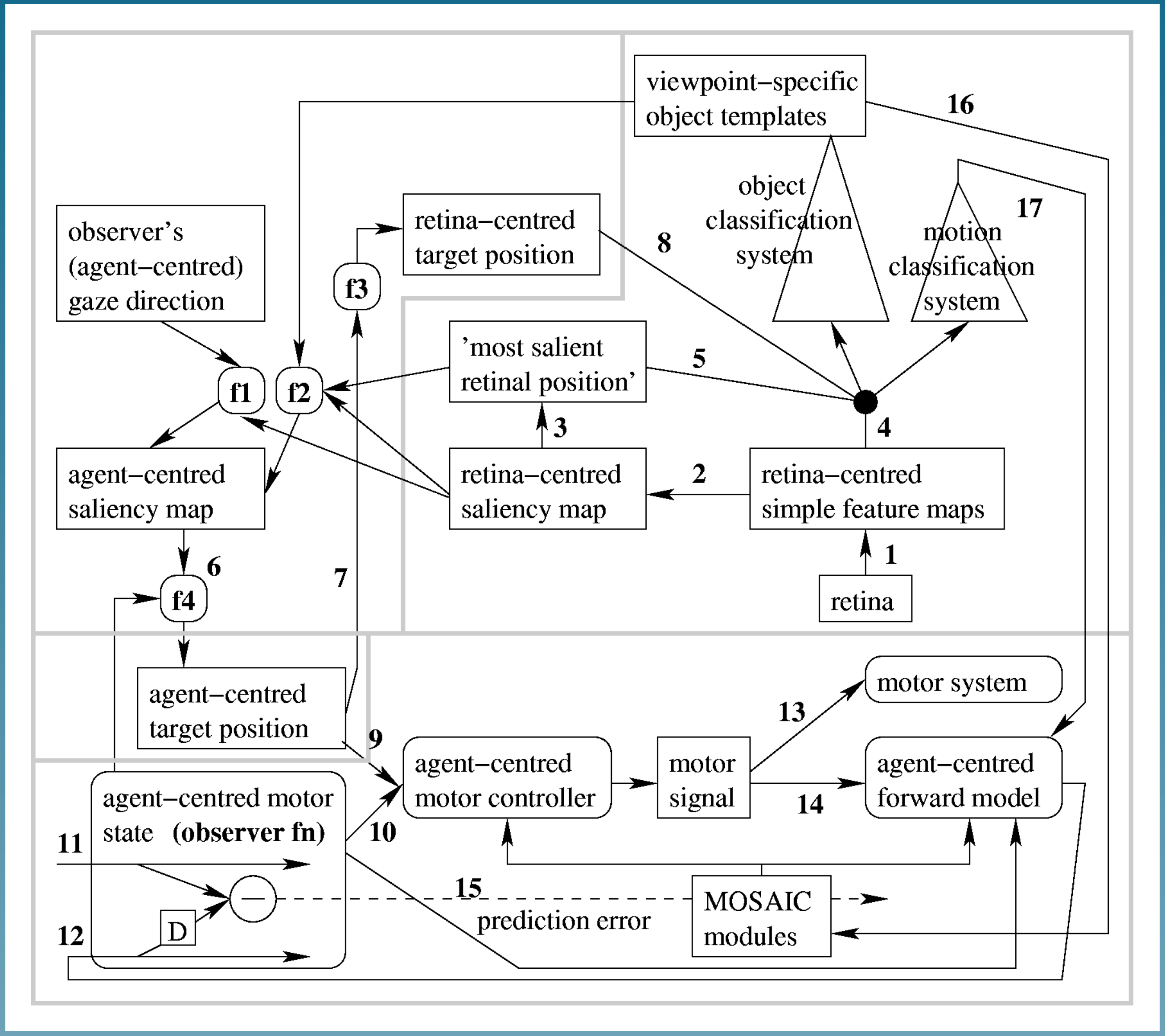
Each cognitive routine creates a deictic representation, which defines (a small amount of) current context.

- An agent needs to learn a **decision function**, giving the best deictic routine to execute in each context.
- Whitehead and Ballard (1990) describe a reinforcement learning architecture for this task. Their system learns to execute a mixture of attentional actions and motor actions.

Action recognition as a deictic routine

1. Start state: a **saliency map** of the current scene.
2. Attention to a known object in this map (the man).
3. Creation of a new saliency map centred on the agent.
4. Attention to the most salient object in this map.
5. Computation of its template and motor affordances.
6. Execution of the action afforded (‘grab’).
7. End state: haptic establishment of the cup by the agent.

Some details:



Episodic memory for deictic routines

When we perceive or execute a grab action, we need to remember it.

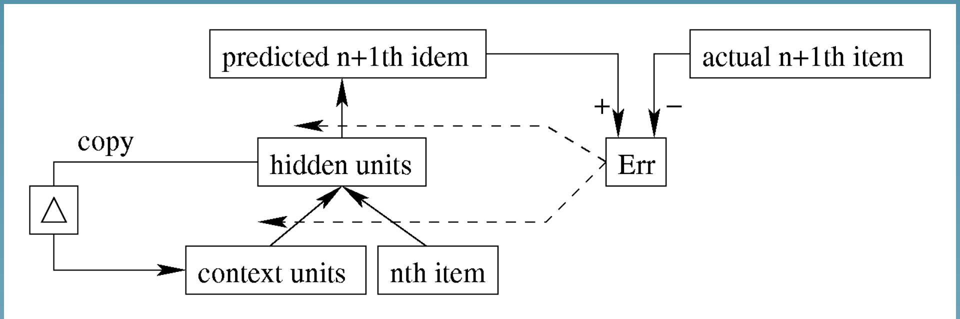
A suggestion:

- What we remember is a sequence of deictic operations, interleaved with their results.
- We encode this sequence using an Elman-style simple recurrent network (SRN).

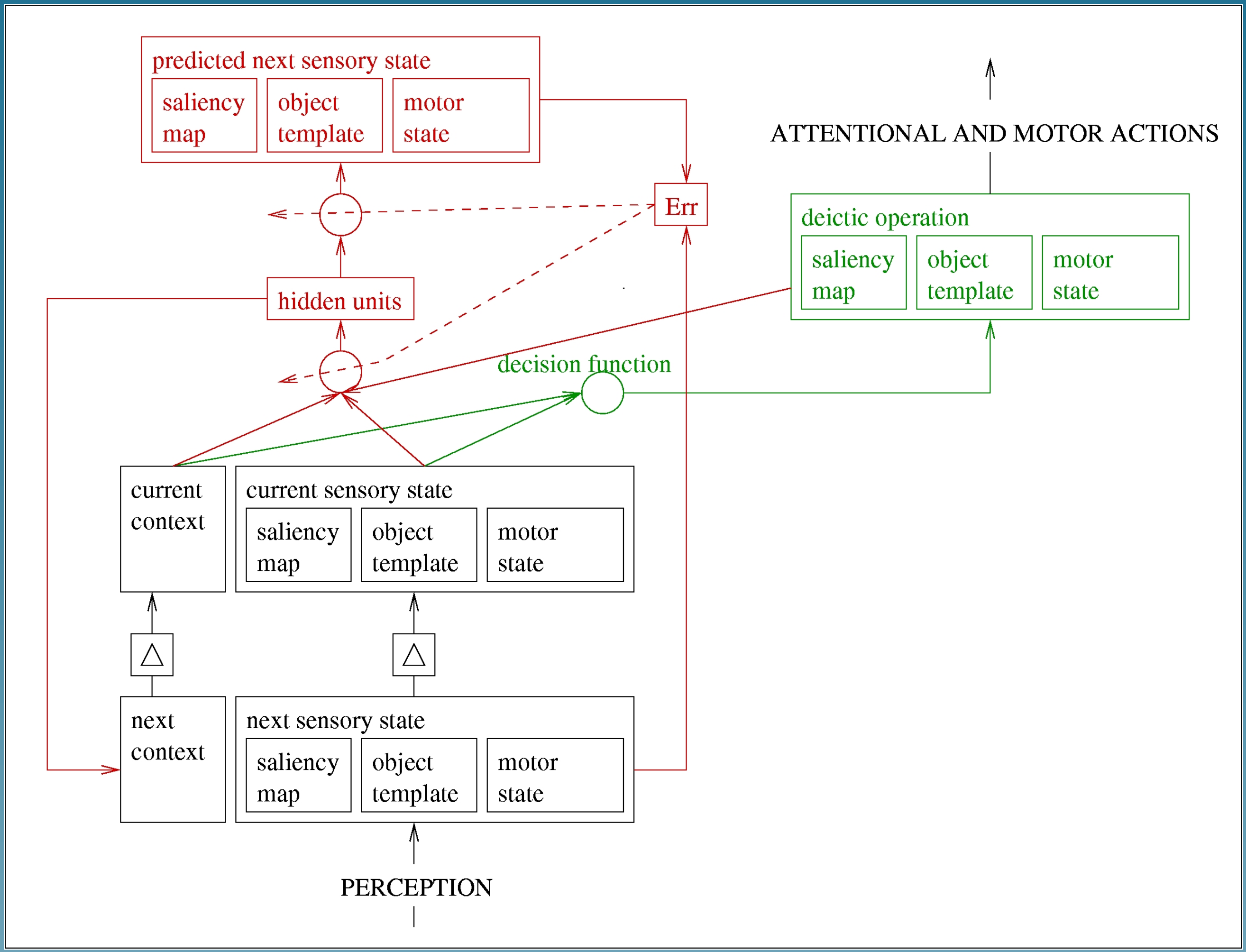
A Simple Recurrent Network

A SRN is presented with a temporal sequence of items as input.

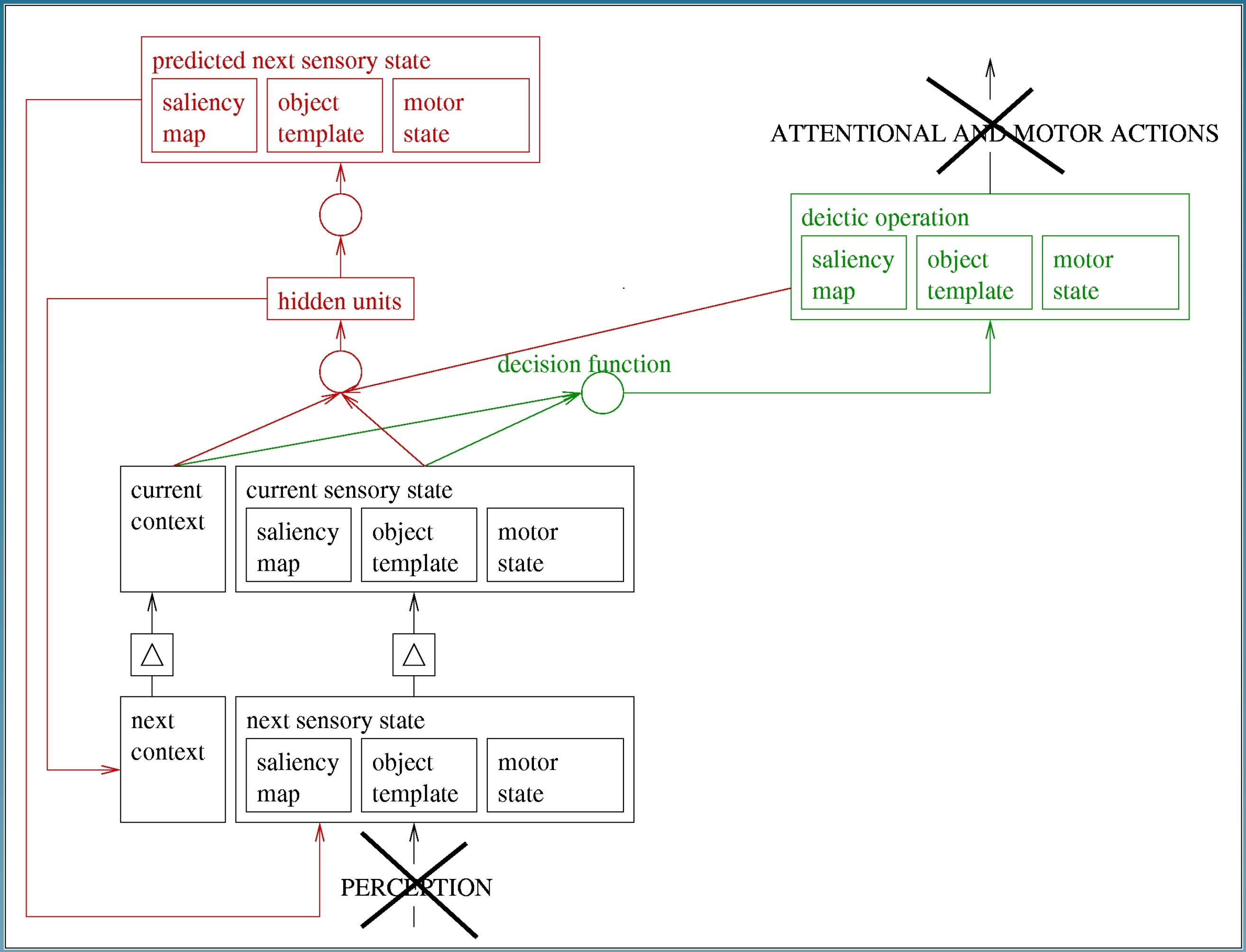
- In training, it is given the next item in the sequence as target output, and it learns to predict this item.
- It solves the task by using a copy of its hidden layer from the previous time point as an additional input.



SRN for deictic routines: ‘experience’ mode



SRN for deictic routines: 'recall' mode



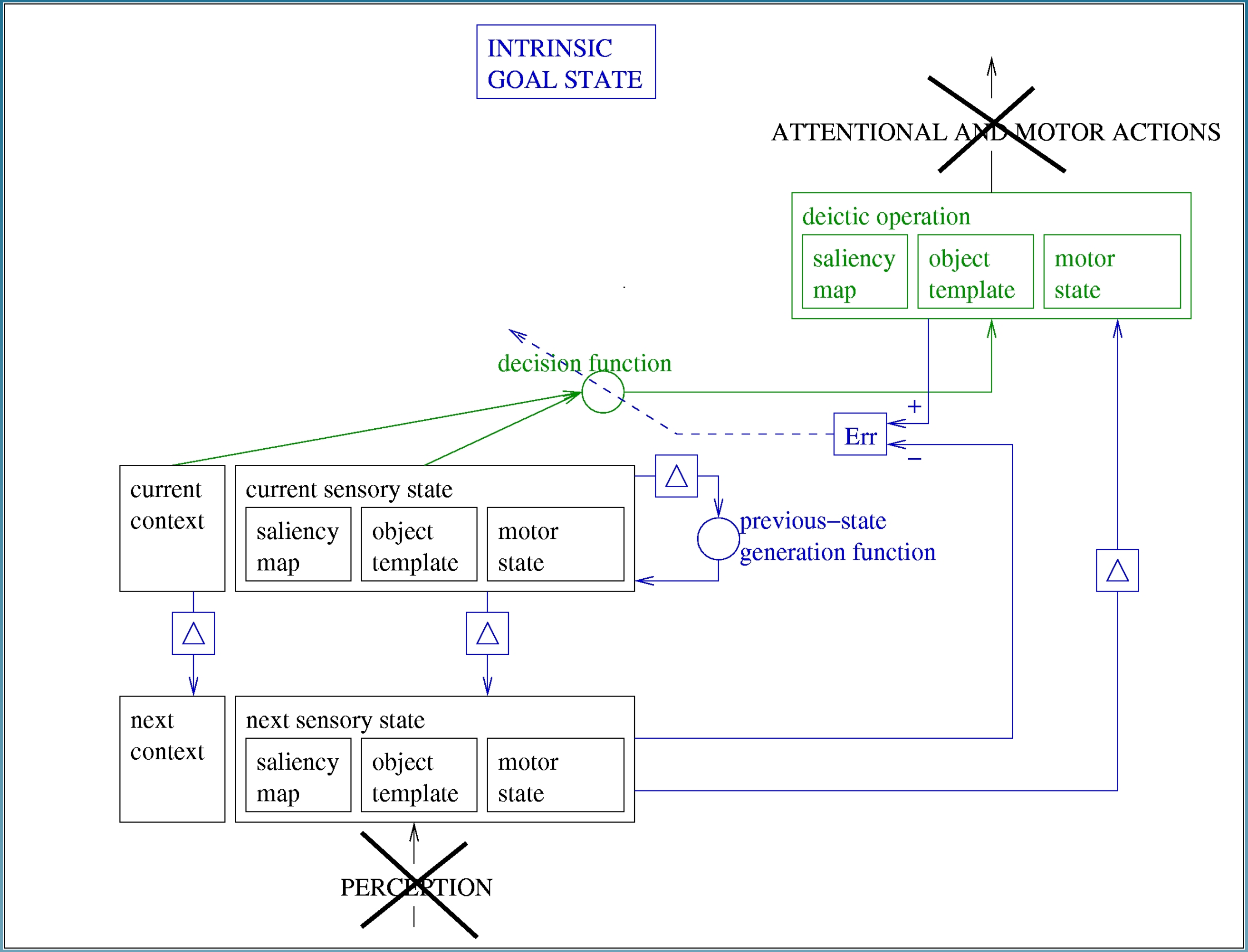
Learning the decision function

If we reach a state/context with **intrinsic value**, we enter a mode in which the decision function is trained.

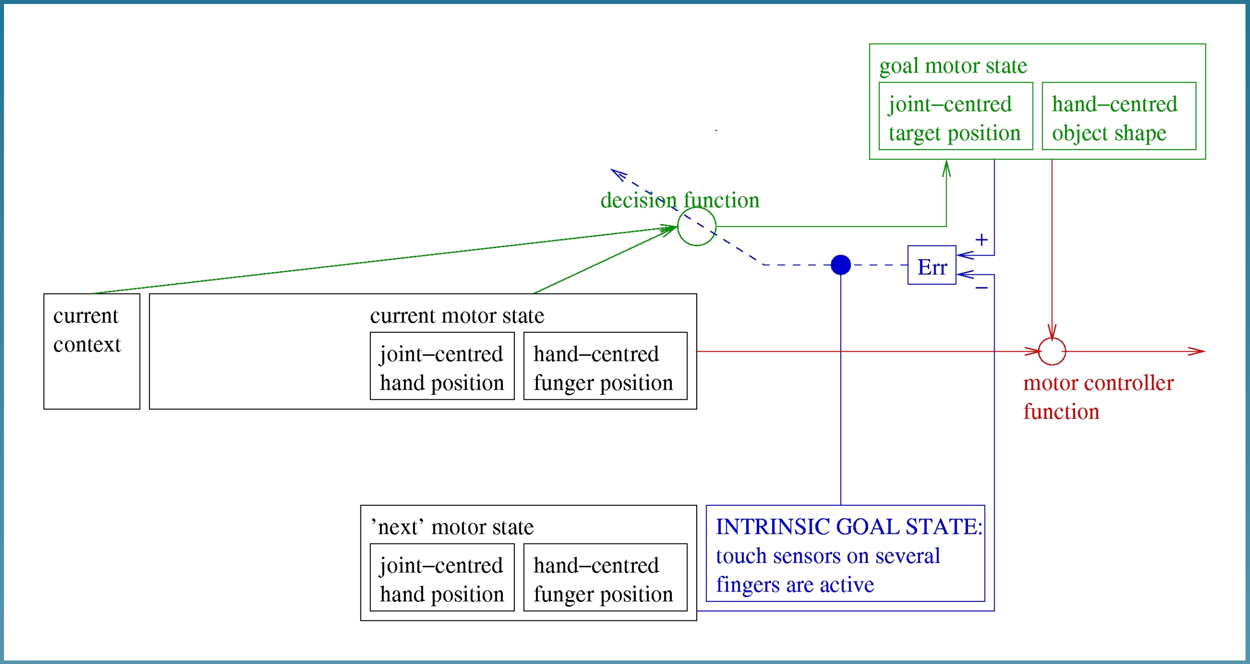
- Key idea: *deictic operations are simply desired states*.
- The training output for the decision function is thus simply the next state—the one with intrinsic value.

To solve the temporal credit assignment problem, the network then iteratively **rewinds**, to train the decision function at earlier points in the sequence.

SRN for deictic routines: ‘rewind’ mode

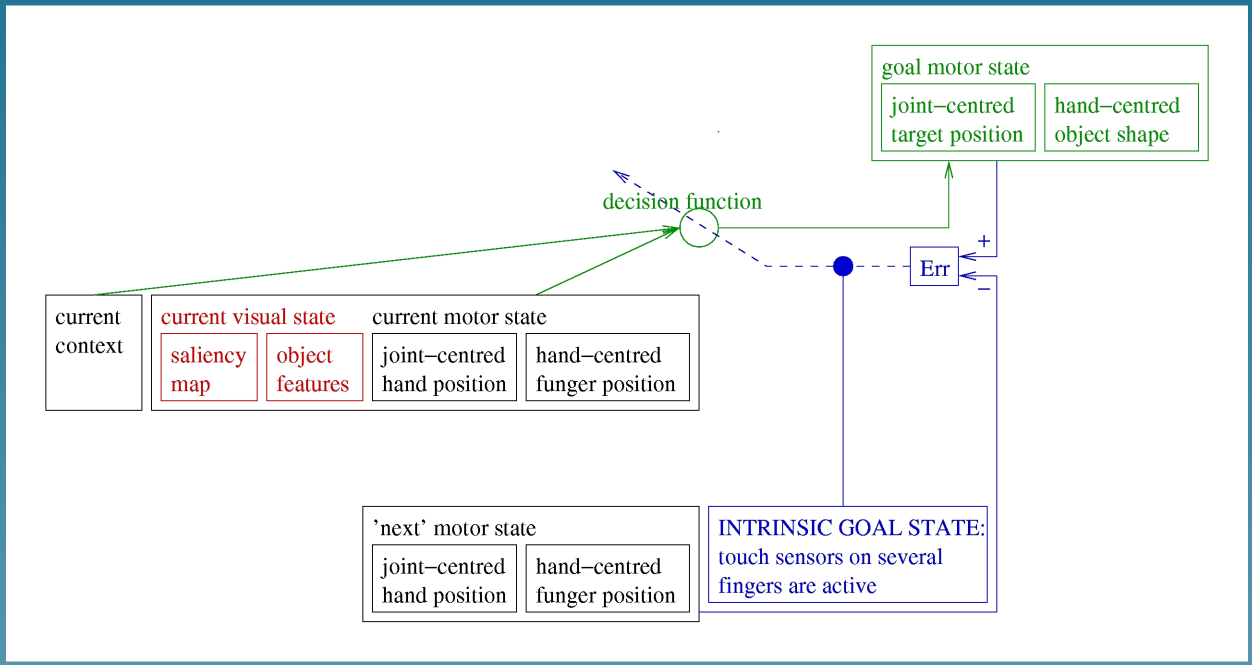


Example: learning a deictic motor operation



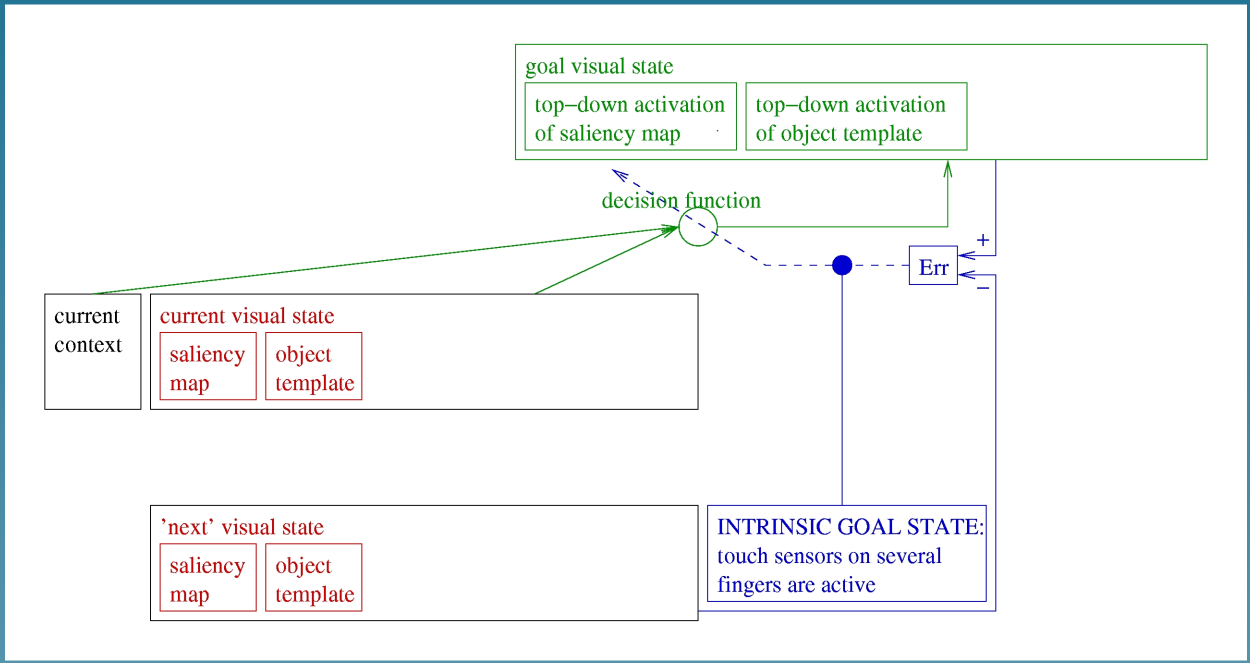
Recall: the **motor controller function** takes input from the current motor state and the goal motor state.

Example: learning sensorimotor mappings



Recall that the input to the decision function includes the current *visual* representations of the target object.

Example: learning an attentional action

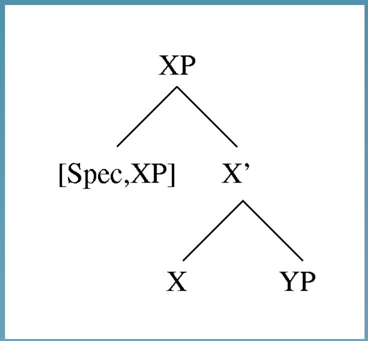


An attentional action is simply top-down activation of the desired object template, and of its expected location.

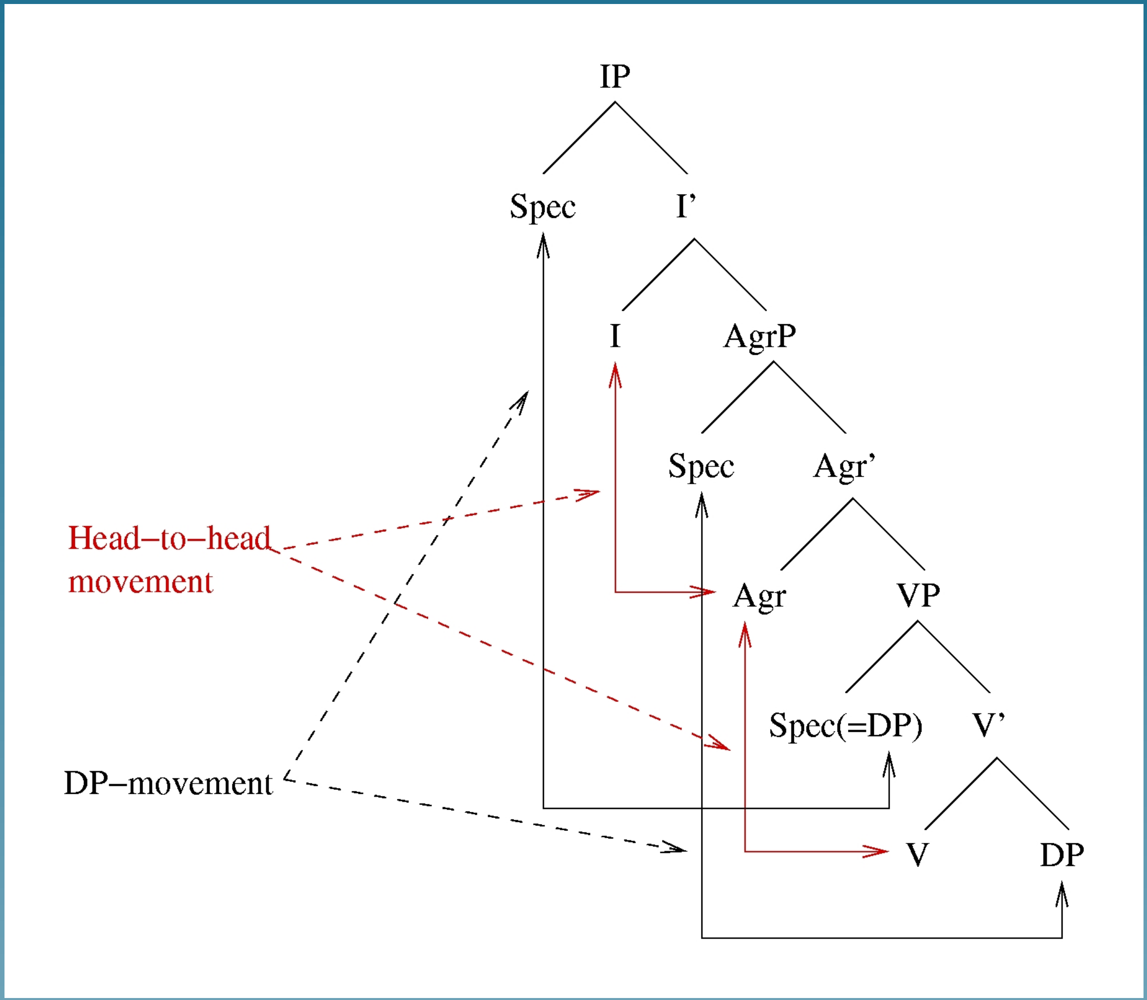
A model of sentence syntax

The syntactic framework I'm using is GB, with a few additions from Minimalism. Very briefly:

- Sentences have a **surface structure** (**SS**) and an underlying **deep structure** (**DS**).
- DS is a series of applications of the **X-bar schema**.



The structure of a transitive clause



Some features of the analysis

- The highest projection is the **inflection phrase** (**IP**).
- Next down is an **agreement phrase** (AgrP), originally posited by Pollock (1989).
- Subject and object DPs receive **theta roles** in [Spec,VP] and [DP,V'] respectively.
- They need to **move** to get **Case** in [Spec,IP] and [Spec,AgrP] respectively.

Head movement and clause finiteness

In finite clauses, V undergoes **head-to-head movement**, raising successively to Agr and then to I. (Or I lowers to V, by the same route.)

(1) *L'homme prend une tasse.*

(2) *The man takes a cup.*

In nonfinite clauses, the verb has no inflection.

(3) *L'homme veut prendre une tasse.*

(4) *The man wants to take a cup.*

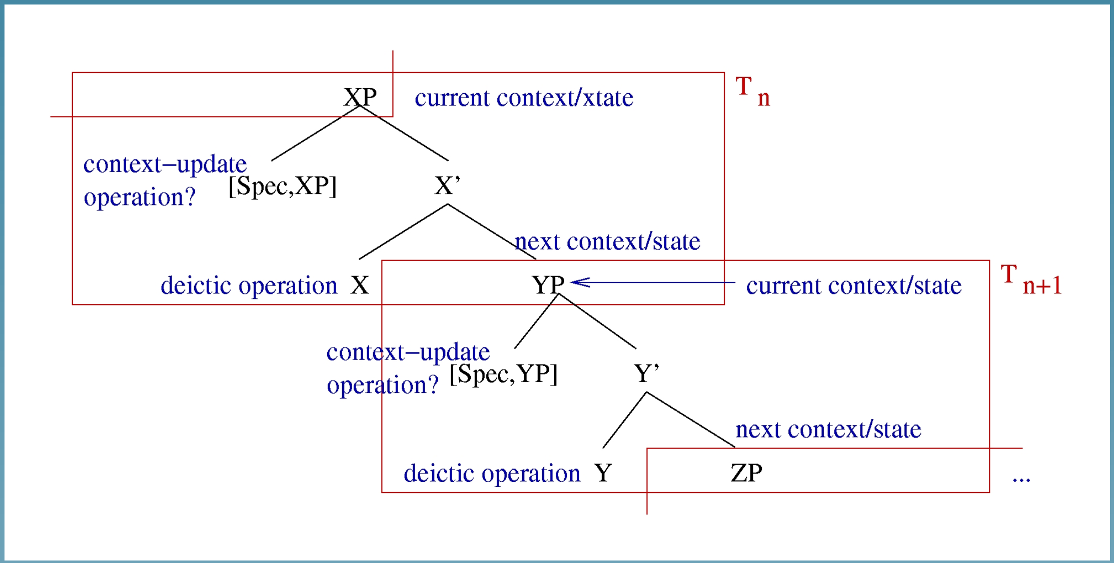
A sensorimotor interpretation of DS

A sensorimotor interpretation of DS

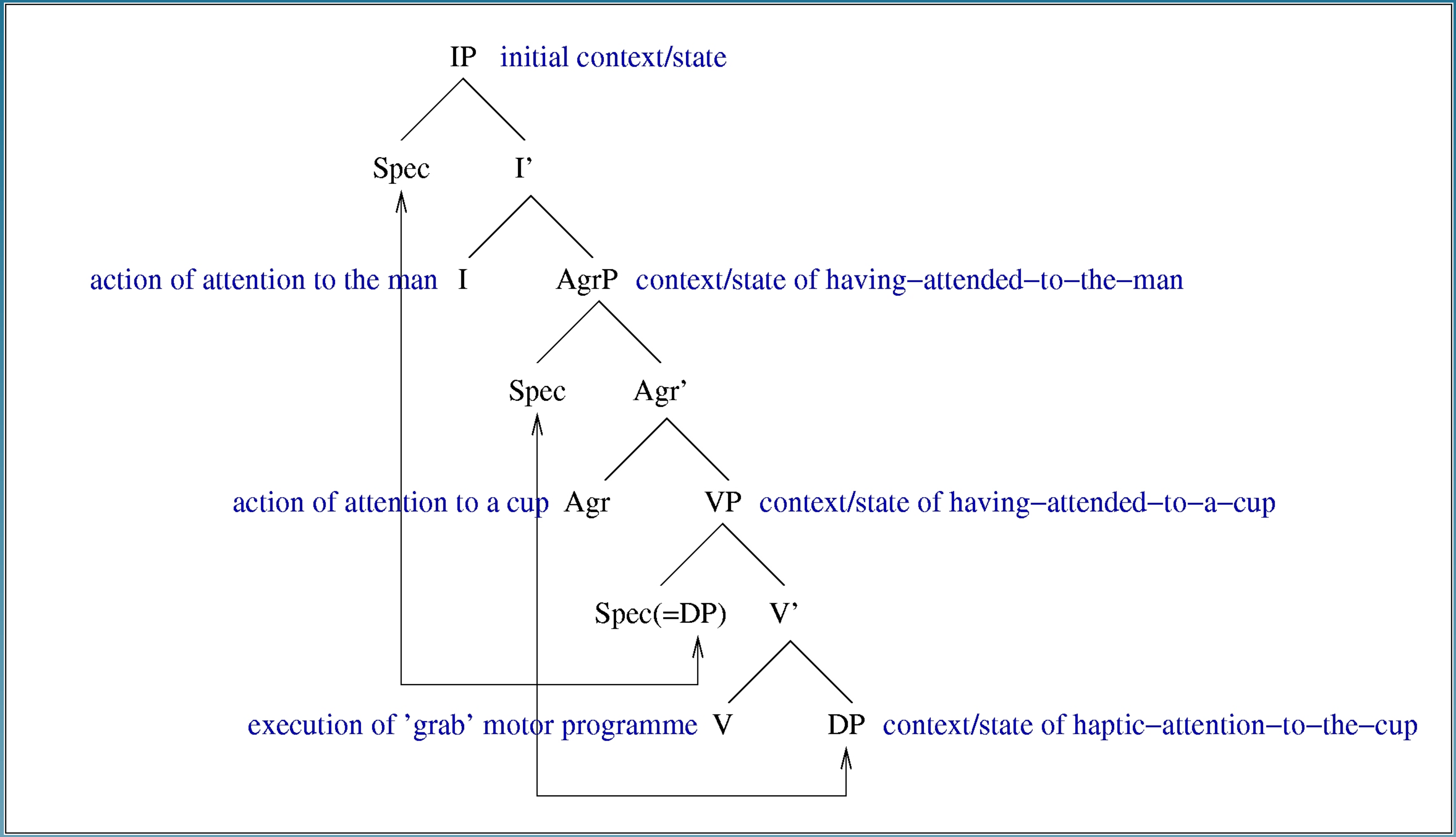
Main idea: the right-branching X-bar structure of DS is an encoding of the representations featuring in the sensorimotor algorithm as it moves from state to state.

A sensorimotor interpretation of DS

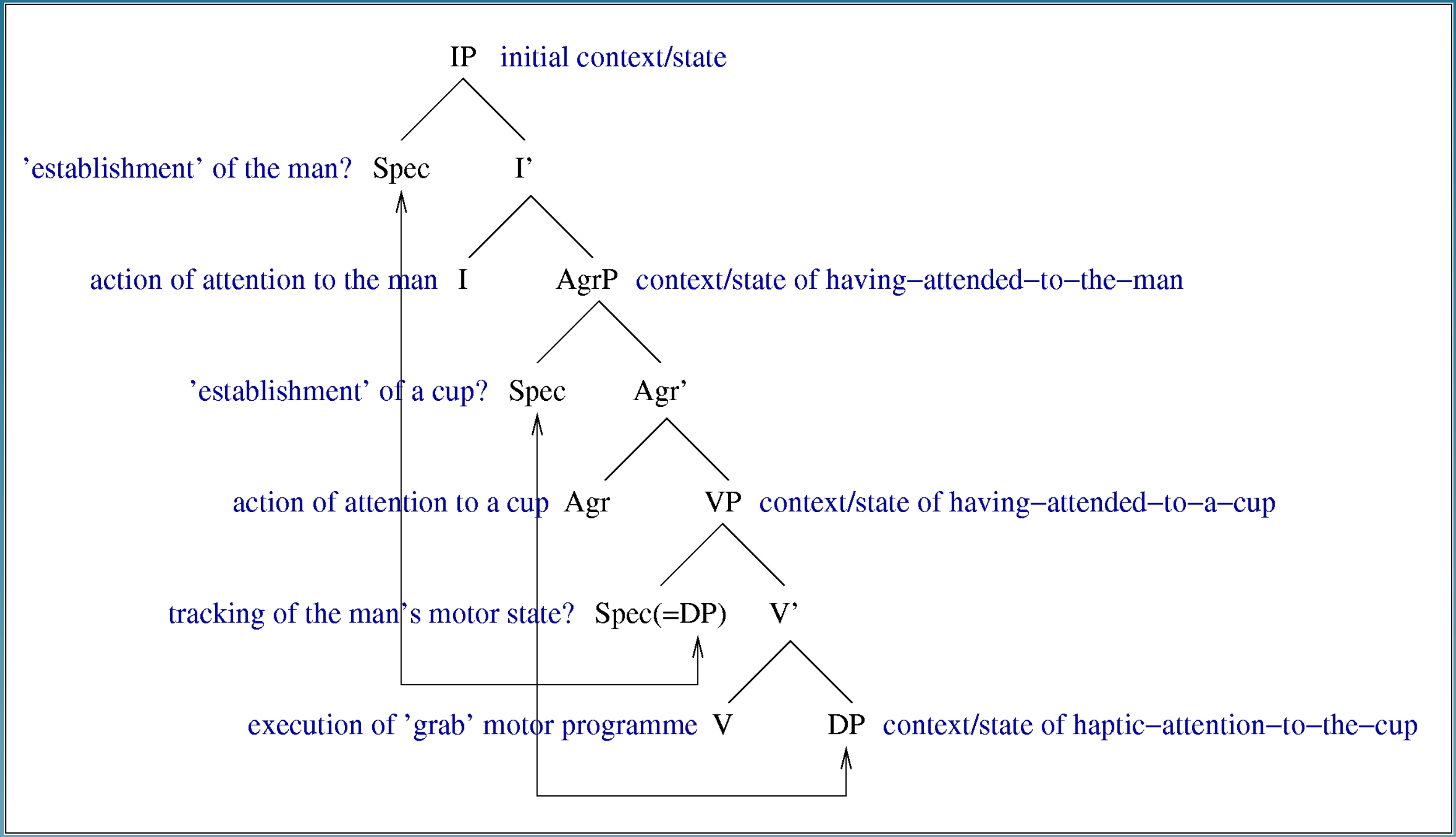
Main idea: the right-branching X-bar structure of DS is an encoding of the representations featuring in the sensorimotor algorithm as it moves from state to state.



Sensorimotor account of a transitive clause



Sensorimotor account of a transitive clause



The mapping from DS to word sequences

Hypothesis: speakers learn to map traces of the sensorimotor algorithm onto sentences by being given *pairs* of traces and sentences.

- Current idea: modify Chang's (1995) 'dual-path' network for predicting the next word in a sentence.
- Constraints on word order will be partly due to the structure of the 'message', and partly to conventions.

Reentrancy and DP-movement

Agent and patient representations both show up more than once in a trace of the algorithm.

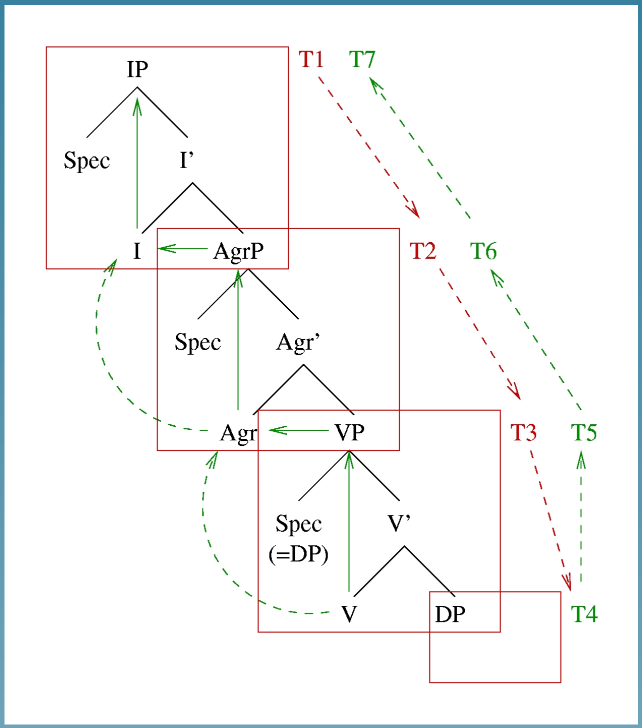
- Agent shows up as a static object ([Spec,IP]) and as a tracked motor state ([Spec,VP]).
- Patient shows up as a static object ([Spec,AgrP]) and as a motor affordance ([DP,V']).

These reentrancies may provide an explanation for DP-movement.

Rewind mode and head-to-head movement

Hypothesis: the V$\rightarrow$Agr$\rightarrow$I movement found in finite sentences encodes the copy operations involved in rewind mode.

- T5: V $\rightarrow$ VP; then DP $\rightarrow$ V.
- T6: Agr $\rightarrow$ AgrP; then VP $\rightarrow$ Agr.
- T7: I $\rightarrow$ IP; then AgrP $\rightarrow$ I.



Conclusions

There may be a formal similarity between a GB-style model of syntax and an independently-motivated model of sensorimotor cognition.

Some future work:

- Finish off the syntactic and sensorimotor models.
- Implement the sensorimotor model (for a simple action).
- Add a learnable mapping from sensorimotor operations to linguistic side-effects.