Syntactic structures as traces of sensorimotor event representations

Alistair Knott
Dept of Computer Science
University of Otago

Two approaches to the study of actions

A simple example event: **a man grabs a cup**.

This event can be studied from two points of view:

- Linguists are interested in analysing sentences which describe the event (e.g. *The man grabbed a cup*).
- Psychologists are interested in understanding how the event can be recognised, performed, remembered etc.

The language/sensorimotor interface

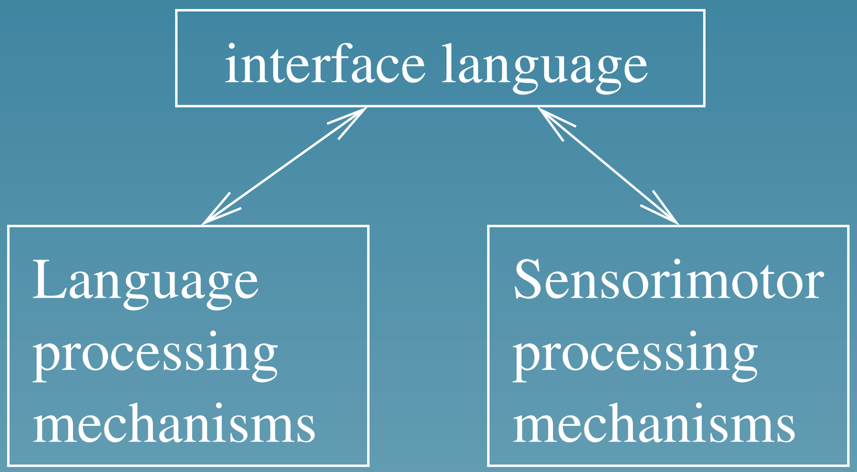
Clearly we can *convert* sensorimotor representations into linguistic ones (and vice versa).

- Because we can talk about what we see;
- Because we can execute verbal instructions.

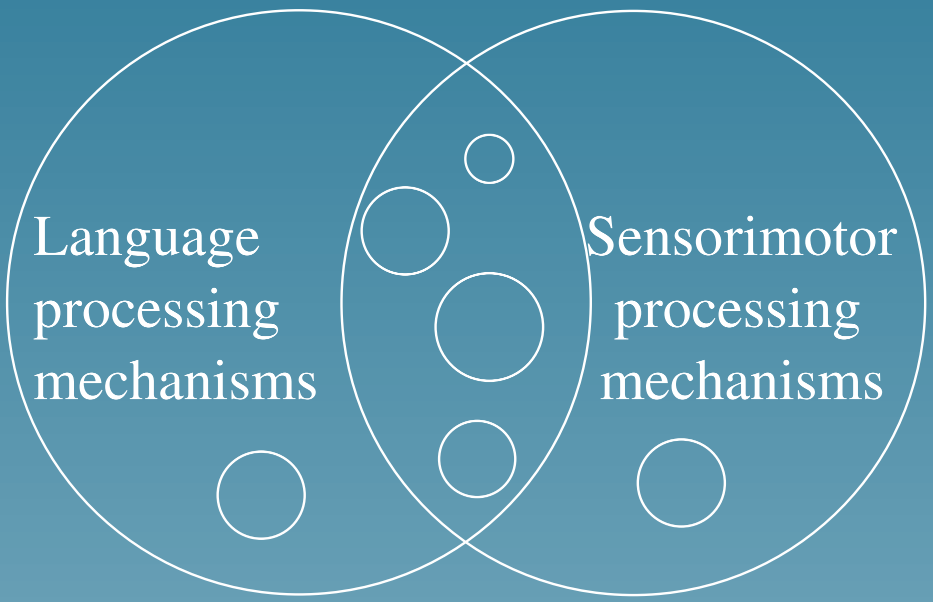
Question: how much work is involved in this conversion?

Two suggestions

Language and SM processing are **modules**



Language and SM processing **share mechanisms**



Methodology

I'm interested in exploring the second suggestion.

My approach is to look for formal similarities between models of sensorimotor cognition and models of sentence syntax.

- If there are nontrivial similarities, then maybe linguists and SM psychologists are actually studying the same thing without realising it.

Outline of the talk

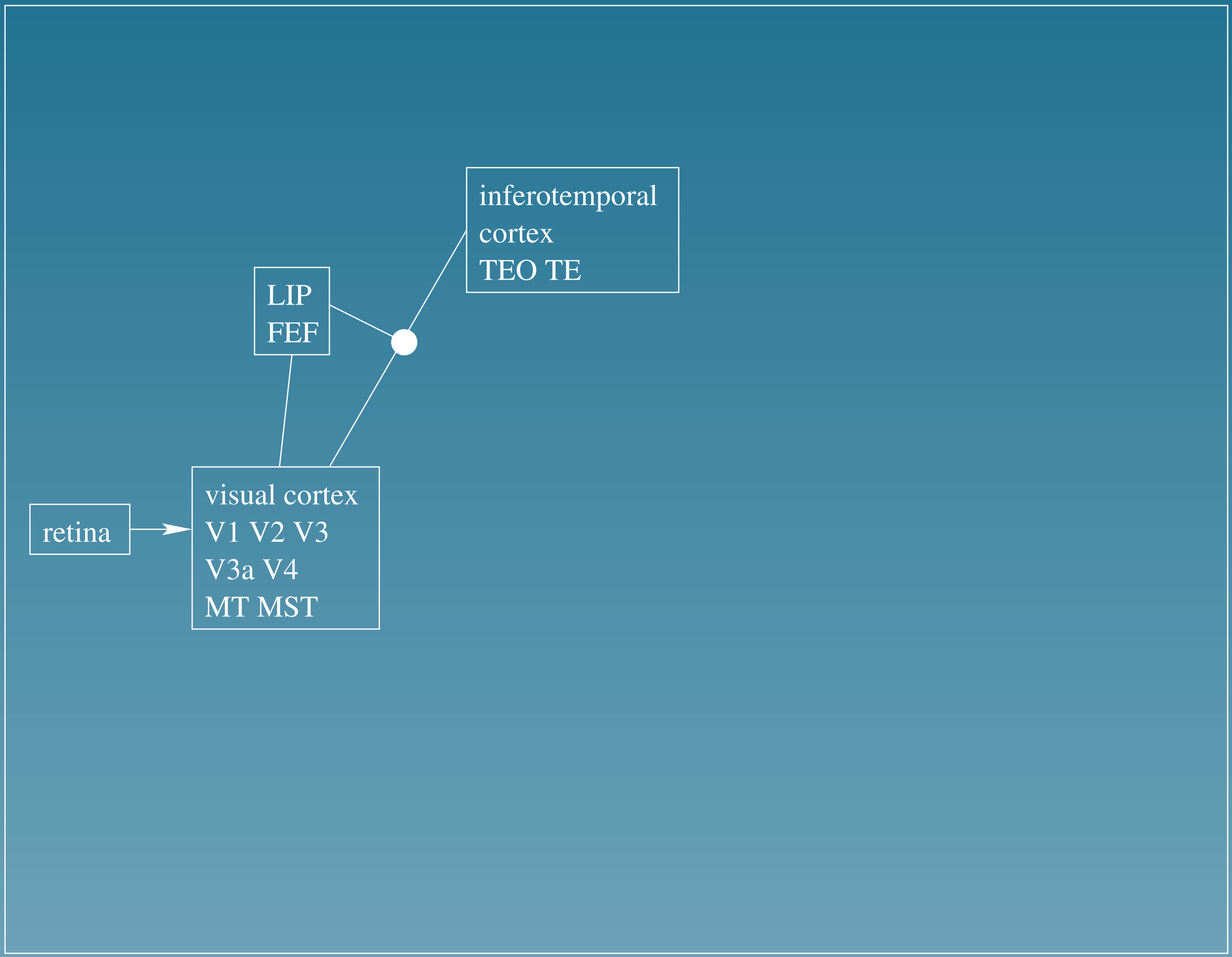
1. A sensorimotor model of transitive actions
2. A syntactic model of transitive actions
3. A suggestion: the syntactic model can be understood as a description of operations in the sensorimotor model.

Preliminaries for the sensorimotor model

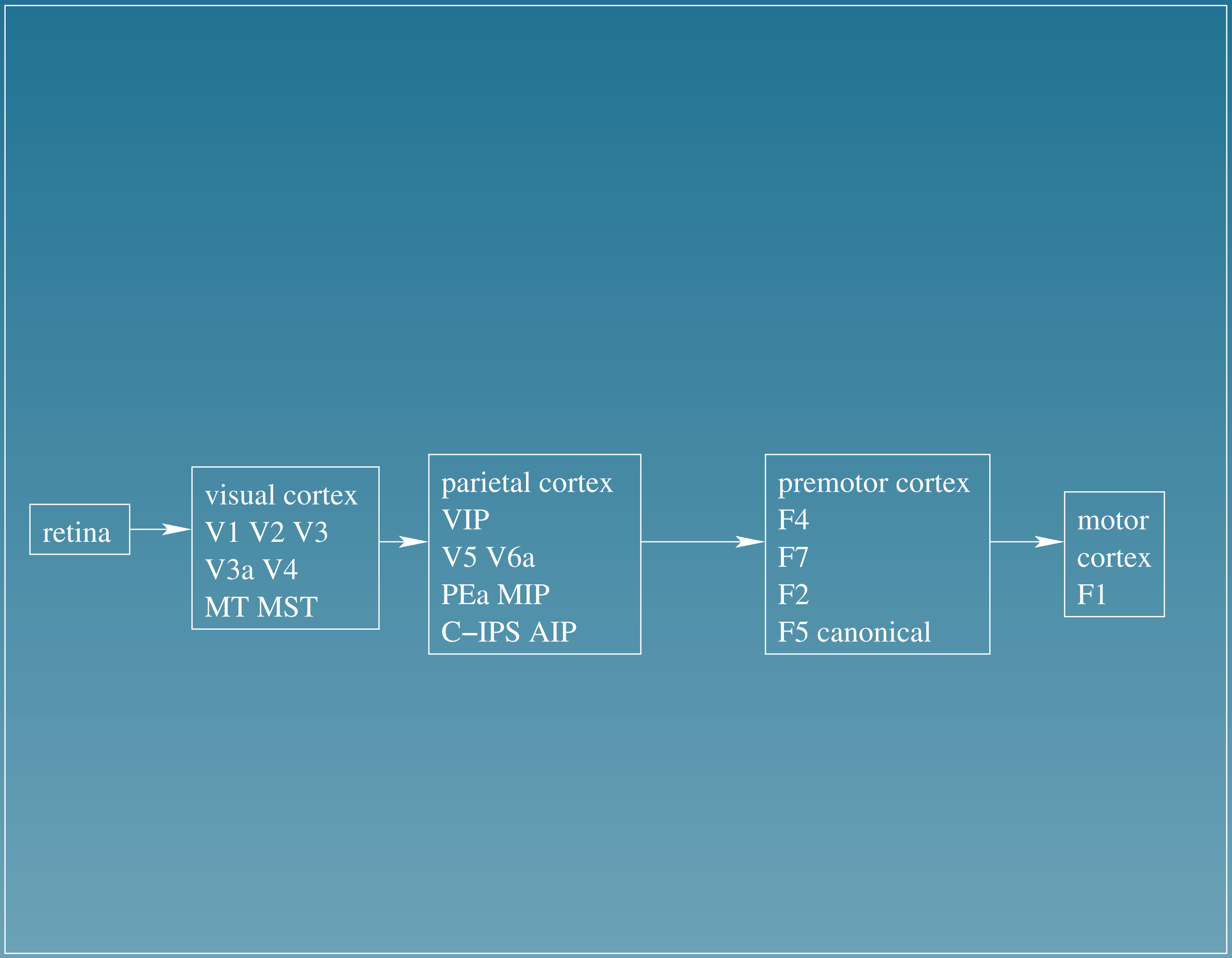
A model of ‘proposition-sized’ cognitive phenomena must draw on work in several different areas of psychology: vision, attention, motor control, working memory, episodic memory.

The model which follows is a synthesis of models from these different areas.

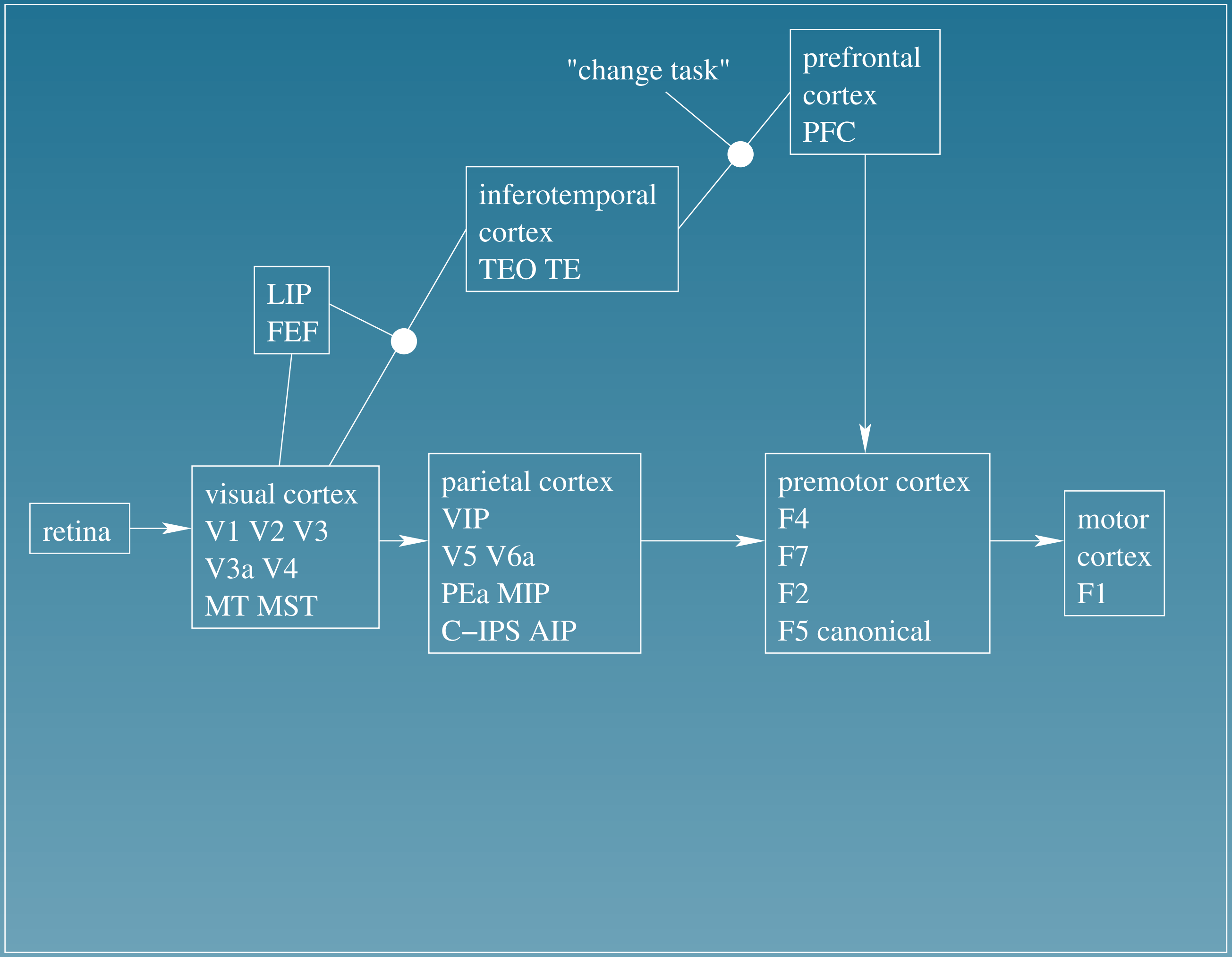
Visual 'what' and 'where' pathways



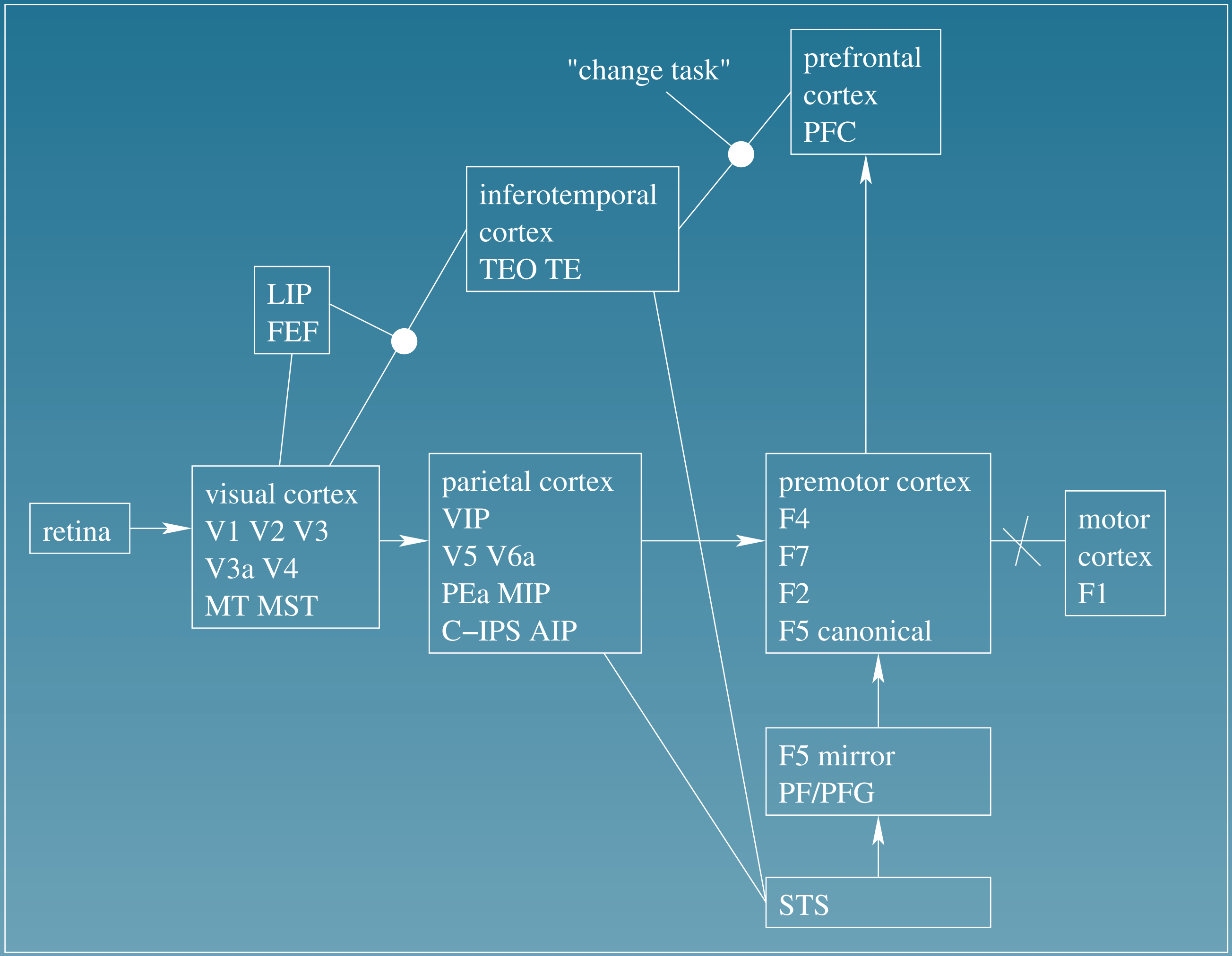
The reach and grasp pathways



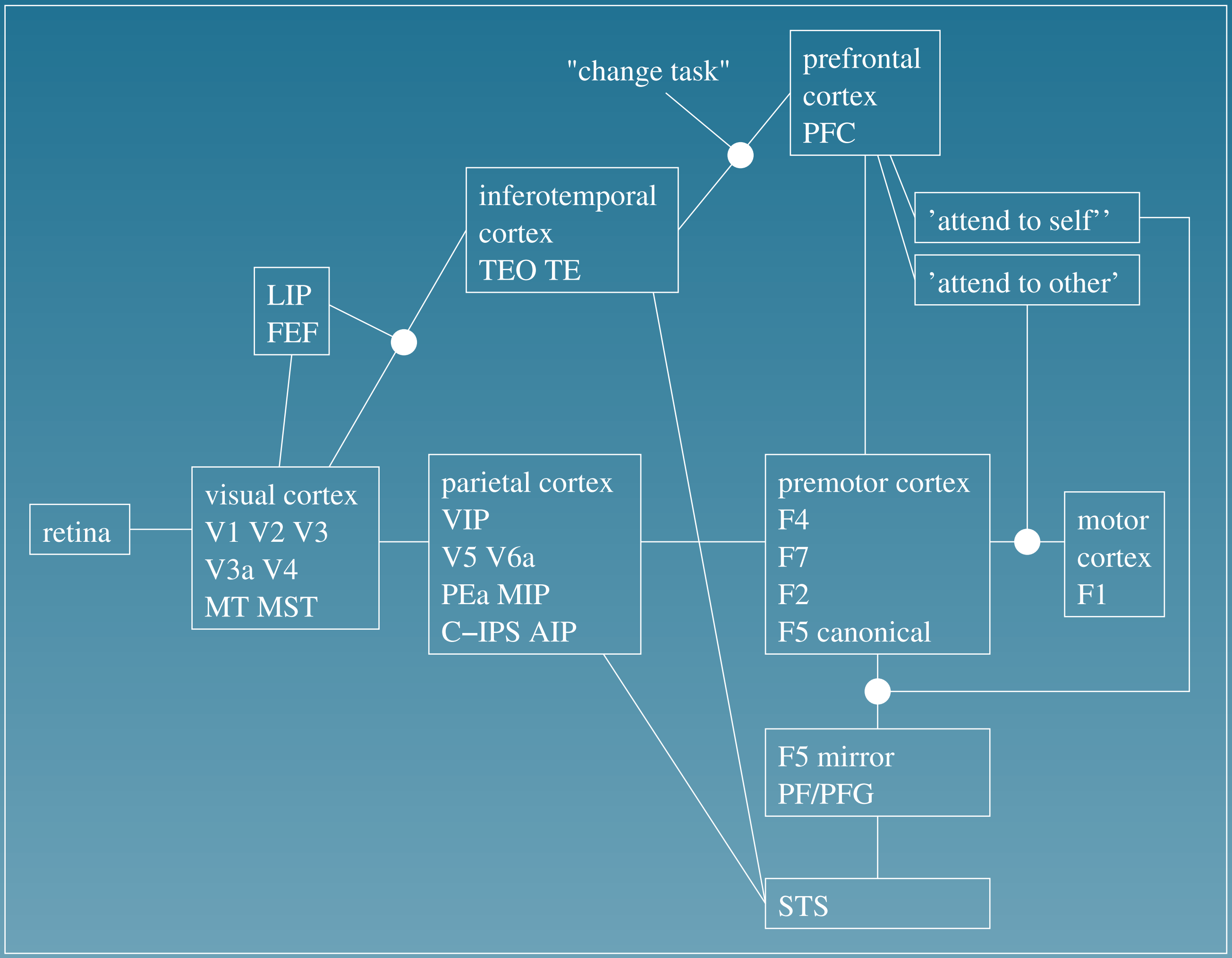
The modulatory role of PFC



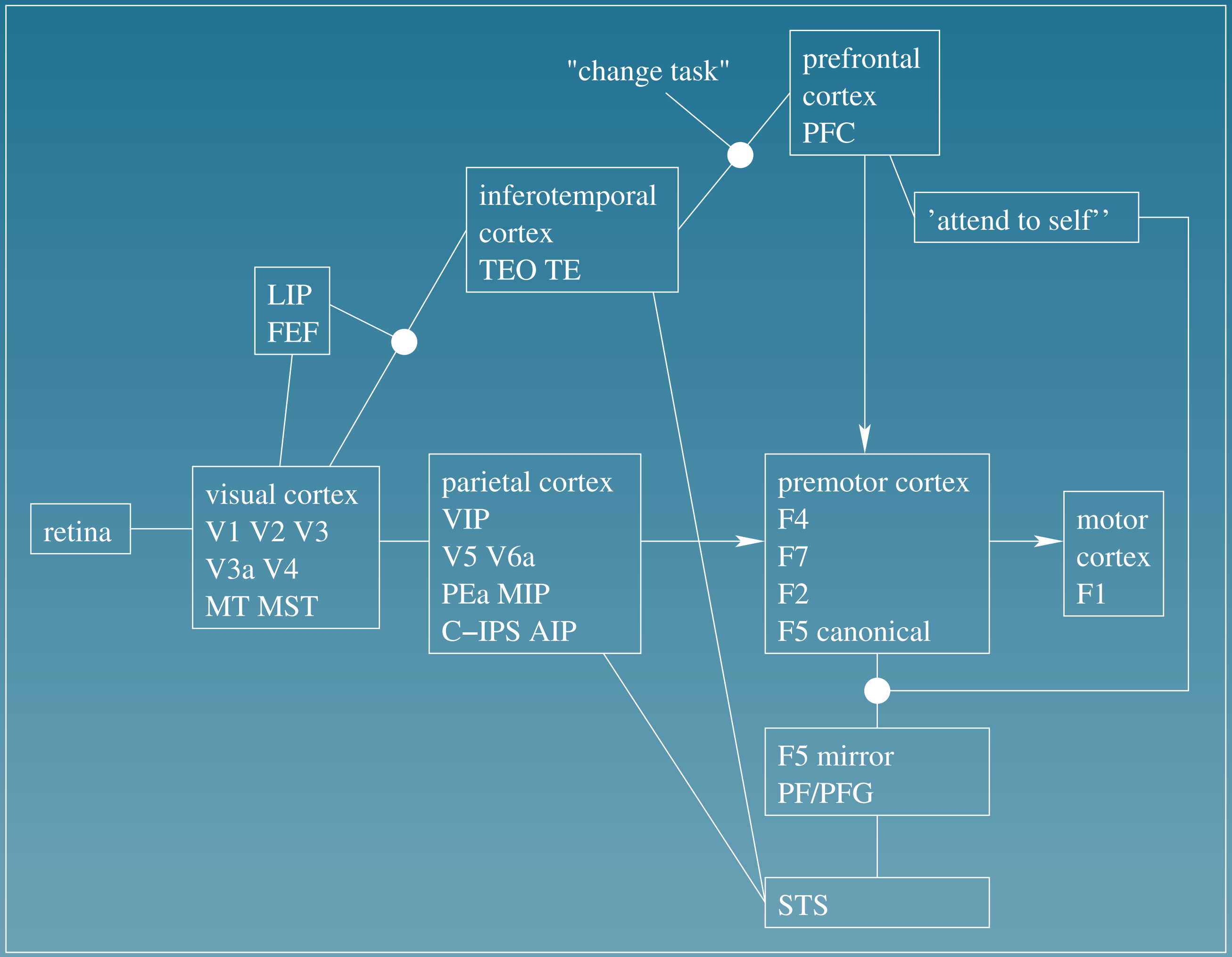
The action recognition pathway



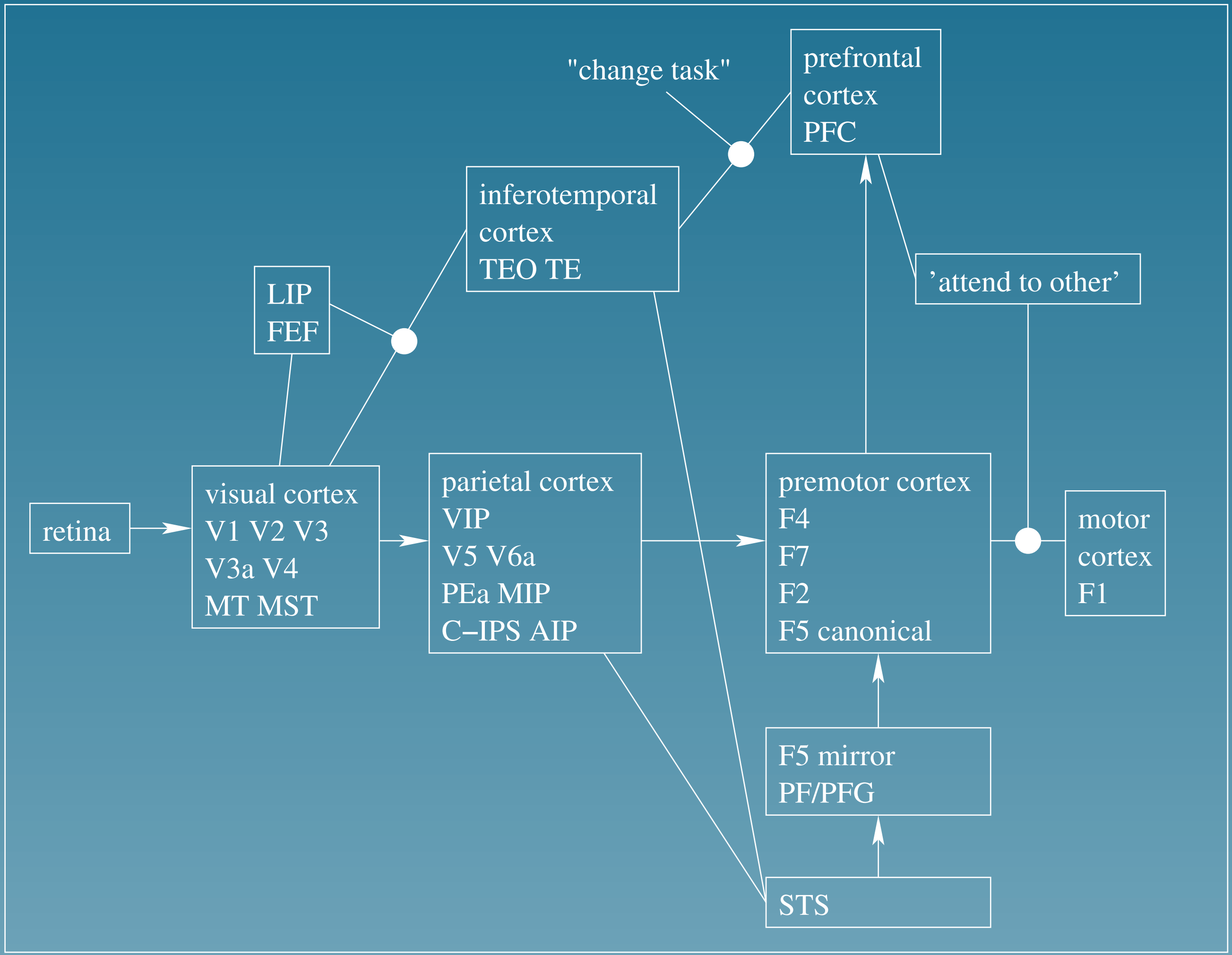
The ‘who’ pathway



‘Action execution mode’



‘Action recognition mode’



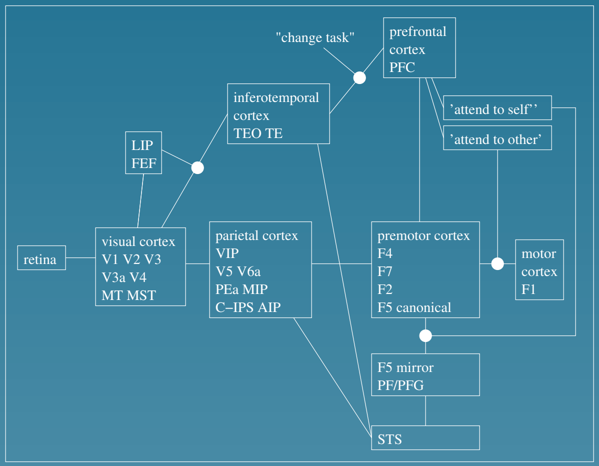
Some references

- IT cortex: object classification (e.g. Logothetis, 1998)
- FEF: saliency map (e.g. Thomson *et al*, 2001)
- Parietal cortex: coordinate system transformations (e.g. Andersen *et al.*, 1997; Burnod *et al.* 1999)
- F4: arm goal position (e.g. Luppino and Rizzolatti, 2000)
- F5: hand motor programmes (e.g. Rizzolatti *et al.*, 1988)
- PFC: Miller and Cohen (2001); Braver and Cohen (2000)

Some references

- Superior temporal sulcus for biological motion recognition (e.g. Oram and Perrett, 1994)
- Mirror neurons in PF/PFG (e.g. Gallese *et al.*, 2002)
- Hebbian models of ‘deep’ action recognition (e.g. Iacoboni *et al.*, 2001; Fogassi *et al.*, 2005; Keysers and Perrett, 2004)
- ‘Mode-setting’ model of self vs other (Farrer and Frith, 2002)

Temporal structure of SM processing



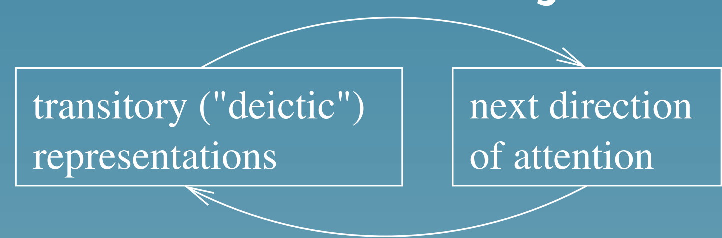
How is processing organised in the network as a whole?

- NB: most representations in the network are *transitory*.
- So it will move through a *sequence* of states.

Deictic routines

A **deictic** representation is a transitory representation linked to the current focus of attention (Ballard *et al.*, 1995).

The current deictic representation can determine how attention is shifted to the *next* object.

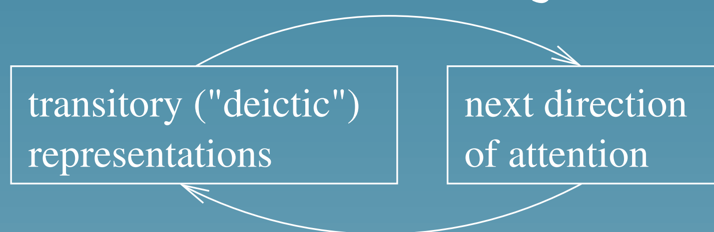


This cycle allows the development of **deictic routines**, involving sequences of directions of attention.

Deictic routines

A **deictic** representation is a transitory representation linked to the current focus of attention (Ballard *et al.*, 1995).

The current deictic representation can determine how attention is shifted to the *next* object.



A proposal: 'events' such as transitive actions are structured as deictic routines.

Proposal about transitive action execution

Stage 1: the observer is in an attentional state where objects in the world compete for his attention.

Stage 2: the observer attends to himself, configuring his mirror system circuit for action execution.

Stage 3: the observer creates a new attentional environment, centred on his own body, biased to objects within reach.

Stage 4: the observer executes an action of attention and ends up attending to (and categorising) a cup.

Proposal about transitive action execution

Stage 1: the observer is in an attentional state where objects in the world compete for his attention.

Stage 2: the observer attends to himself, configuring his mirror system circuit for action execution.

Stage 3: the observer creates a new attentional environment, centred on his own body, biased to objects within reach.

Stage 4: the observer executes an action of attention and ends up attending to (and categorising) a cup.

Proposal about transitive action execution

Stage 1: the observer is in an attentional state where objects in the world compete for his attention.

Stage 2: the observer attends to himself, configuring his mirror system circuit for action execution.

Stage 3: the observer creates a new attentional environment, centred on his own body, biased to objects within reach.

Stage 4: the observer executes an action of attention and ends up attending to (and categorising) a cup.

Proposal about transitive action execution

Stage 1: the observer is in an attentional state where objects in the world compete for his attention.

Stage 2: the observer attends to himself, configuring his mirror system circuit for action execution.

Stage 3: the observer creates a new attentional environment, centred on his own body, biased to objects within reach.

Stage 4: the observer executes an action of attention and ends up attending to (and categorising) a cup.

Proposal about transitive action execution

Stage 5: the observer creates a new attentional environment, in which several possible alternative actions (on the cup) are represented, and compete amongst one another.

Stage 6: the observer selects one of these actions ('grab'). This triggers physical motion. As a side-effect of this motion, the observer *again* attends to himself.

Stage 7: when the action is completed, the agent *again* attends to the cup, this time in the **haptic** modality.

Proposal about transitive action execution

Stage 5: the observer creates a new attentional environment, in which several possible alternative actions (on the cup) are represented, and compete amongst one another.

Stage 6: the observer selects one of these actions ('grab'). This triggers physical motion. As a side-effect of this motion, the observer *again* attends to himself.

Stage 7: when the action is completed, the agent *again* attends to the cup, this time in the **haptic** modality.

Proposal about transitive action execution

Stage 5: the observer creates a new attentional environment, in which several possible alternative actions (on the cup) are represented, and compete amongst one another.

Stage 6: the observer selects one of these actions ('grab'). This triggers physical motion. As a side-effect of this motion, the observer *again* attends to himself.

Stage 7: when the action is completed, the agent *again* attends to the cup, this time in the **haptic** modality.

Proposal about trans. action recognition

Stage 1: the observer is in an attentional state where objects in the world compete for his attention.

Stage 2: the observer attends to a third party, establishing the 'action recognition' mirror circuit.

Stage 3: the observer creates a new attentional environment, centred on the agent and biased towards objects in his peripersonal space.

Stage 4: before the agent's action is complete, the observer saccades to the target and categorises it.

Proposal about trans. action recognition

Stage 1: the observer is in an attentional state where objects in the world compete for his attention.

Stage 2: the observer attends to a third party, establishing the 'action recognition' mirror circuit.

Stage 3: the observer creates a new attentional environment, centred on the agent and biased towards objects in his peripersonal space.

Stage 4: before the agent's action is complete, the observer saccades to the target and categorises it.

Proposal about trans. action recognition

Stage 1: the observer is in an attentional state where objects in the world compete for his attention.

Stage 2: the observer attends to a third party, establishing the 'action recognition' mirror circuit.

Stage 3: the observer creates a new attentional environment, centred on the agent and biased towards objects in his peripersonal space.

Stage 4: before the agent's action is complete, the observer saccades to the target and categorises it.

Proposal about trans. action recognition

Stage 1: the observer is in an attentional state where objects in the world compete for his attention.

Stage 2: the observer attends to a third party, establishing the 'action recognition' mirror circuit.

Stage 3: the observer creates a new attentional environment, centred on the agent and biased towards objects in his peripersonal space.

Stage 4: before the agent's action is complete, the observer saccades to the target and categorises it.

Proposal about trans. action recognition

Stage 5: the observer creates a new attentional environment, in which several possible alternative actions on the cup compete amongst one another.

Stage 6: the observer's biological motion recognition system is engaged, and the correct action ('grab') is identified. As a side-effect, the observer attends back to the agent.

Stage 7: when the action is completed, the observer again attends to the cup.

Proposal about trans. action recognition

Stage 5: the observer creates a new attentional environment, in which several possible alternative actions on the cup compete amongst one another.

Stage 6: the observer's biological motion recognition system is engaged, and the correct action ('grab') is identified. As a side-effect, the observer attends back to the agent.

Stage 7: when the action is completed, the observer again attends to the cup.

Proposal about trans. action recognition

Stage 5: the observer creates a new attentional environment, in which several possible alternative actions on the cup compete amongst one another.

Stage 6: the observer's biological motion recognition system is engaged, and the correct action ('grab') is identified. As a side-effect, the observer attends back to the agent.

Stage 7: when the action is completed, the observer again attends to the cup.

Support for the model

Attention is needed for object categorisation (Treisman and Gelade, 1980 &ff)	$S1 < S2$ $S3 < S4$
Attention can use a body-centred reference frame (Bisiach, 1986; Andersen, 2000)	S3
Info about the agent's posture is needed to create a body-centred reference frame	$S2 < S3$
Only attended-to targets elicit F5 grasp responses (Rizzolatti *et al*)	$S4 < S5$
Attention can use an action-centred frame of reference (Tipper *et al*, 1998)	S5

Support for the model

Object categorisation occurs during biological motion processing (Giese, 2000)	S6
Biological motion processing requires attention (Cavanagh *et al.*, 2001)	S4 $<$ S6
An object must be attended to before it can be reached for (Jeannerod, 1996)	S4 $<$ S6
The target of an observed action is anticipated by the observer (Flanagan, 2003)	S4 $<$ S5/6

A sequence-based semantics for actions

Summary: a transitive action is perceived as a *sequence*, in which the agent, patient and action occupy characteristic serial positions.

A suggestion: transitive actions are not only *perceived* as sequences, but stored in working memory as such.

- WM can hold *planned* actions (i.e. planned sequences)
- WM can hold *observed actions* (i.e. stored sequences)

PFC and working memory

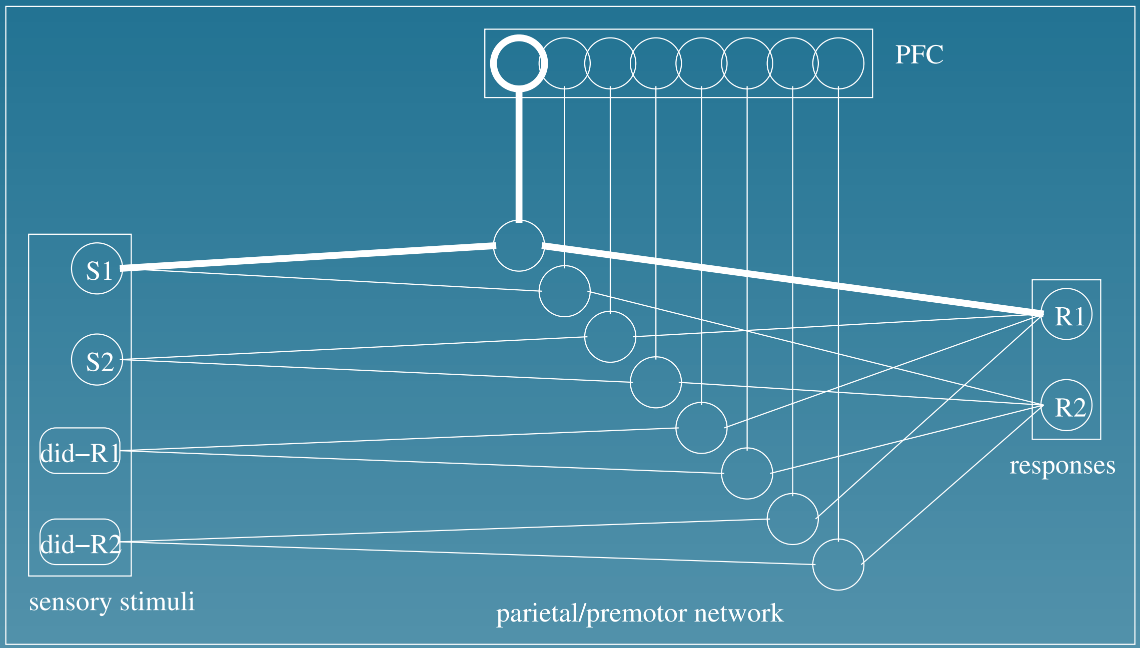
PFC is held to be the locus of many 'working memory' functions.

So, we might imagine that:

- Prior to executing a reach action, the agent activates a PFC representation encoding a planned sequence
- When observing a reach action, the observer *ends up* activating a PFC representation encoding this same plan

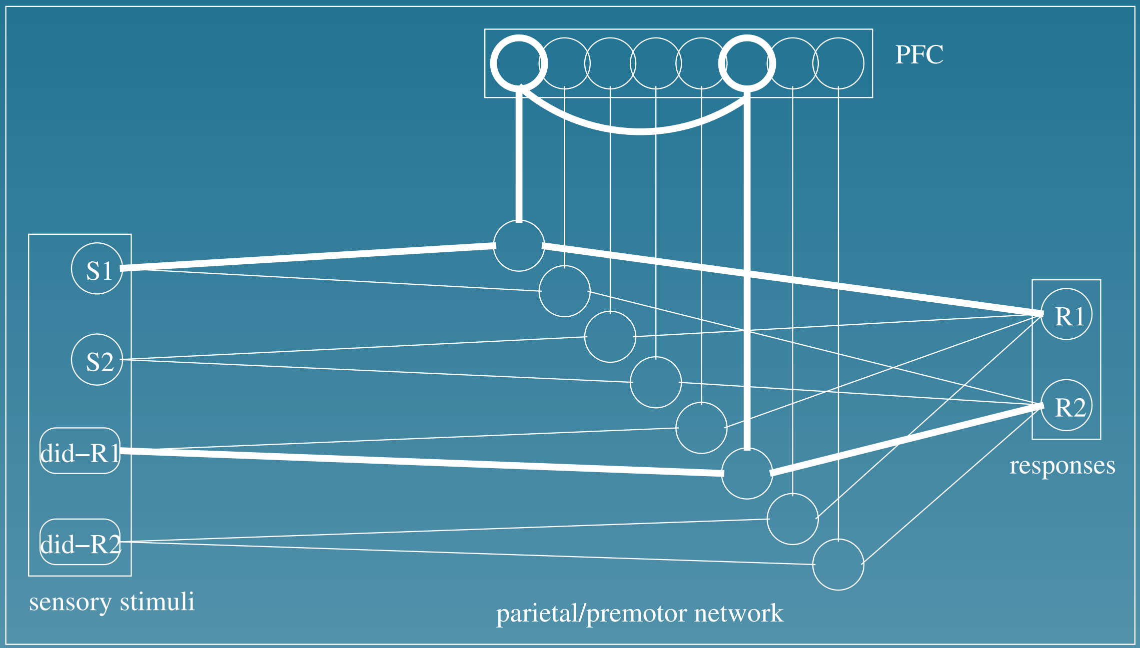
Q: What might this PFC representation look like?

Miller and Cohen's model of PFC



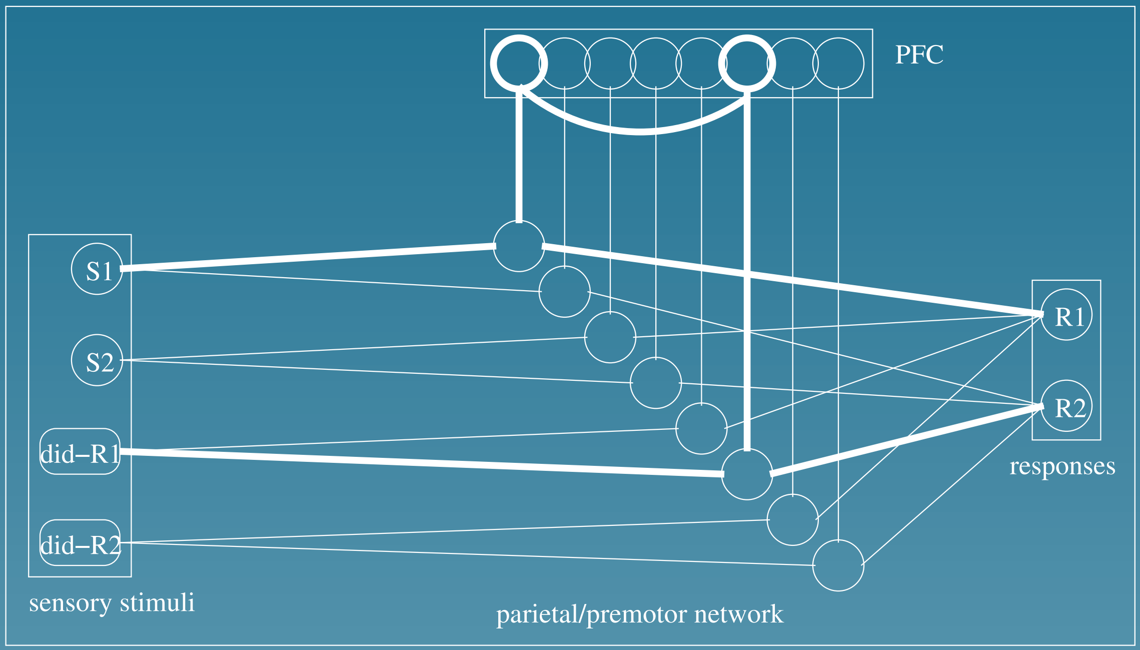
In Miller and Cohen's model, PFC biases neurons in the stimulus-response pathway, influencing competition between them, and selecting particular S-R pathways.

A model of sequence representation in PFC



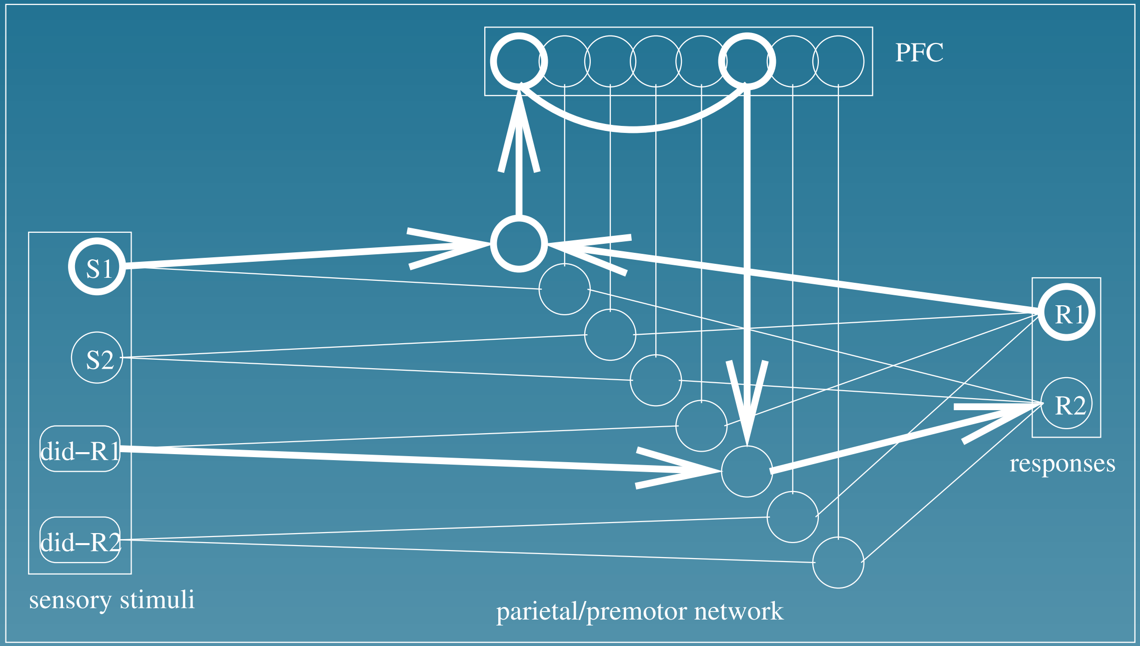
If reafferent sensory consequences of earlier actions count as new sensory stimuli, PFC can represent planned *sequences* of actions.

A model of sequence representation in PFC



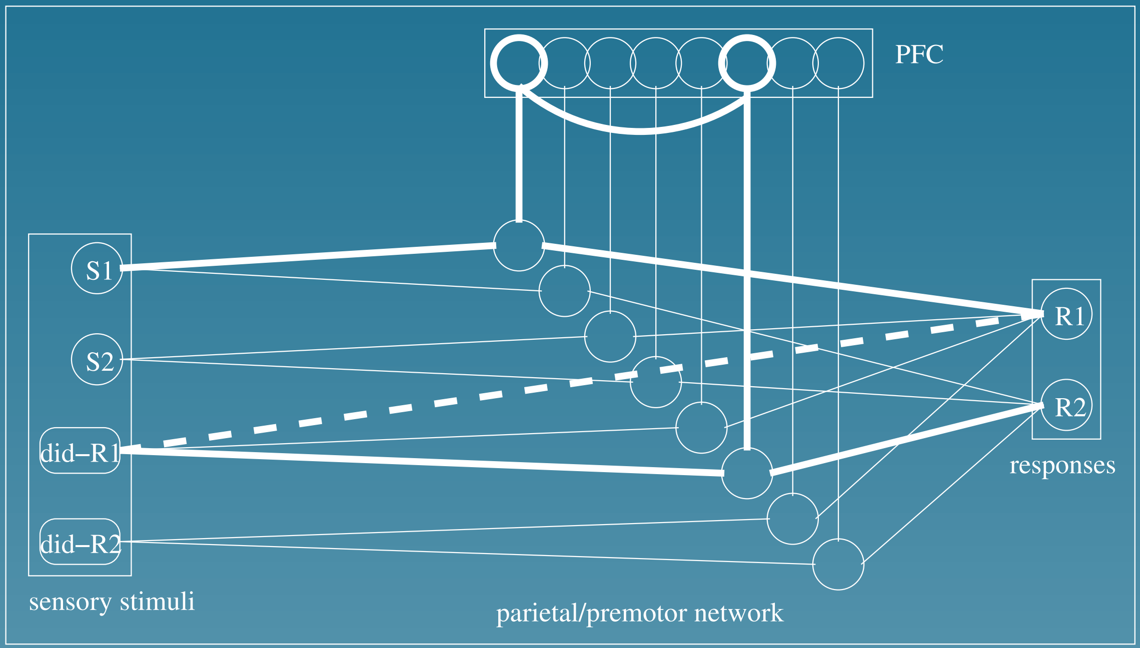
The PFC representation is tonically active before, during and after sequence execution. So it can also operate in action observation, to hold the agent's inferred intention.

A model of sequence representation in PFC



The PFC representation is tonically active before, during and after sequence execution. So it can also operate in action observation, to hold the agent's inferred intention.

A model of sequence representation in PFC



This PFC representation could also support *replaying* of executed or perceived sequences to longer-term memory stores (c.f. Baddeley's 'episodic buffer').

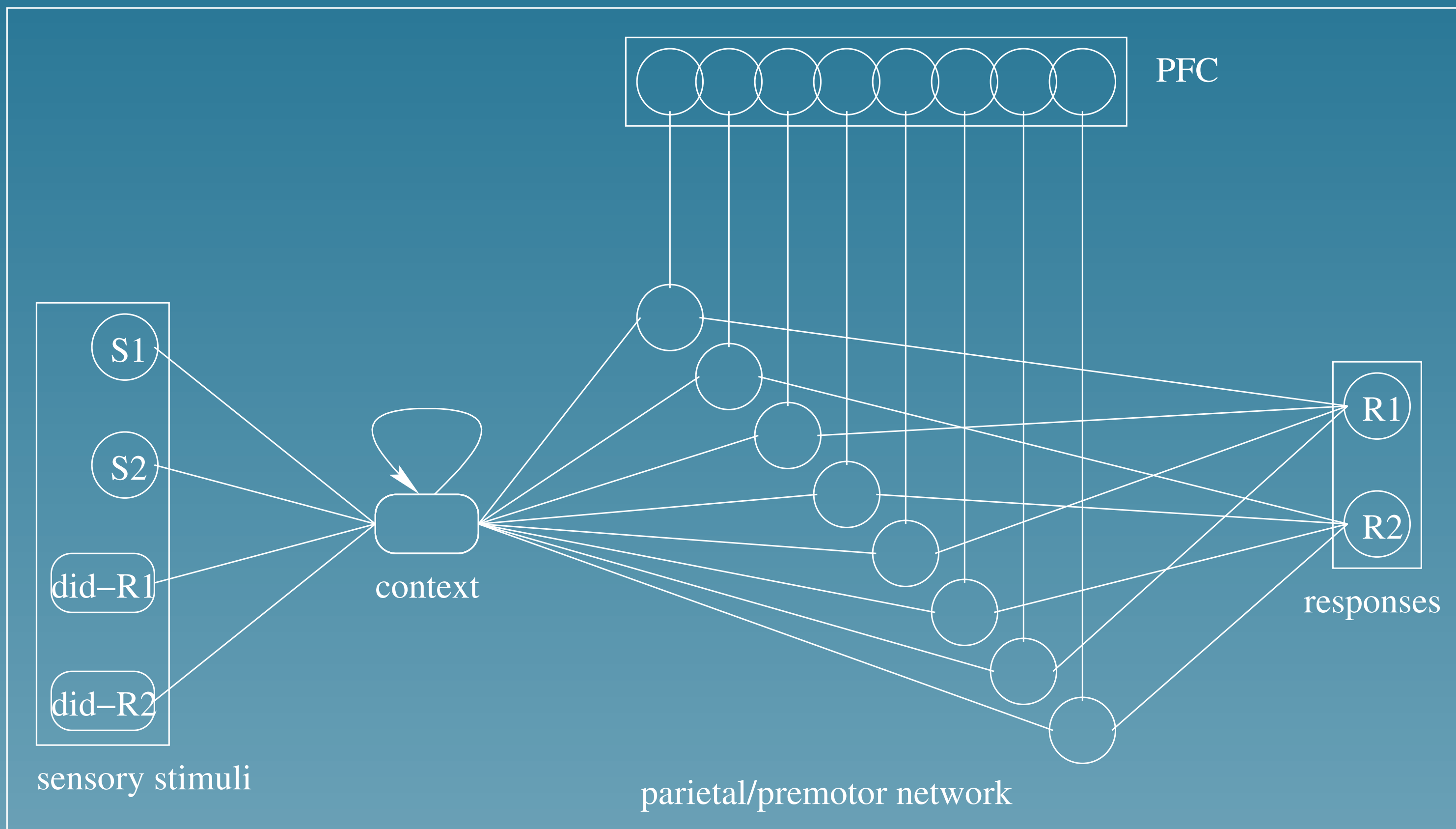
A role for 'context' representions

It's probably unrealistic to assume that the 'input' to the S-R pathway is always a *single* stimulus.

I assume that stimuli have their influence indirectly, by updating a representation of 'the current context'.

- Current context is computed from the most recent stimulus, and also from its previous state.
- Context could be stored in another PFC area (see e.g. Beiser and Houk's model of sequence encoding).

A role for ‘context’ representions



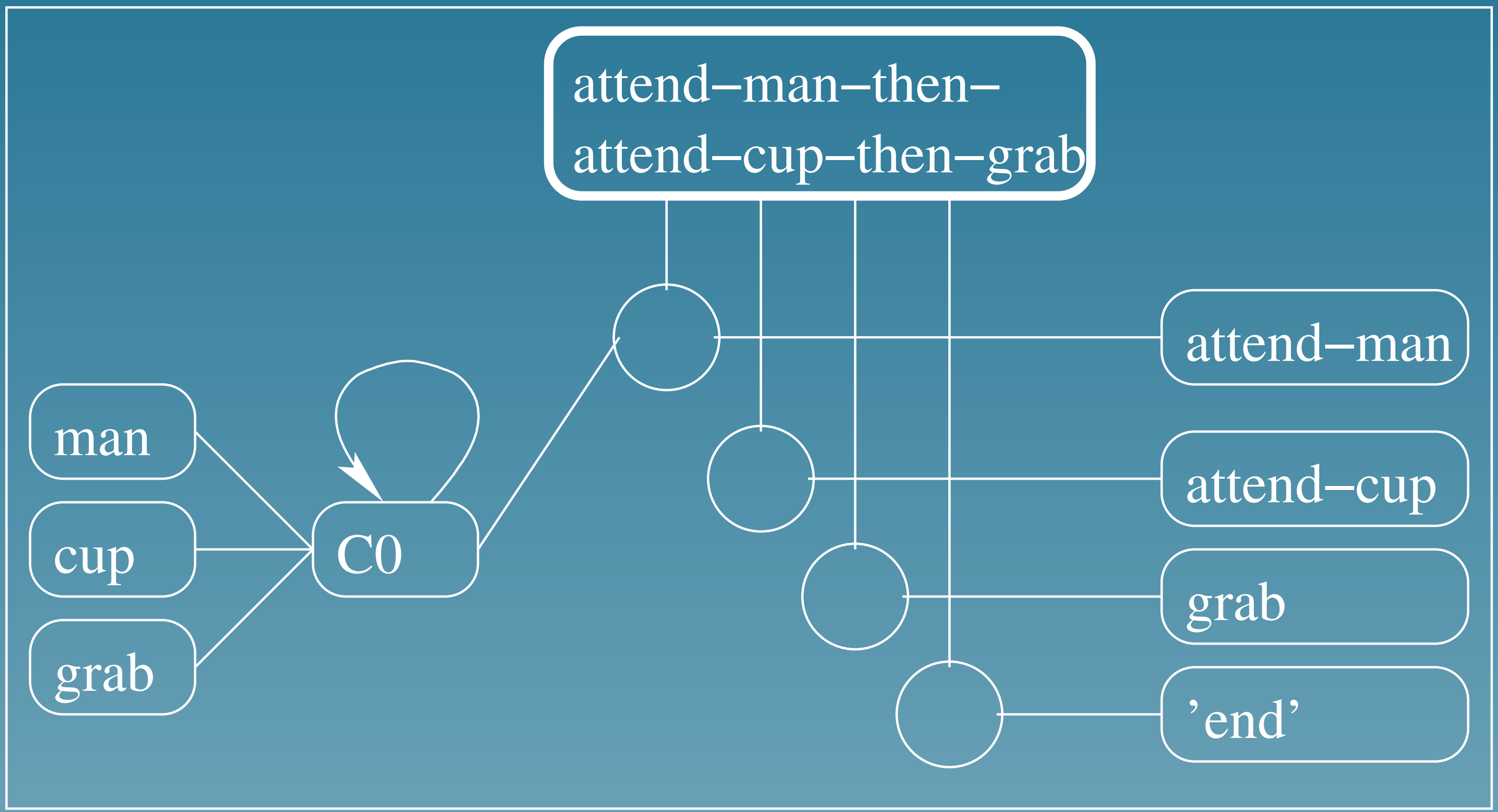
Sensorimotor model: summary

So: what's the SM representation of a transitive action?

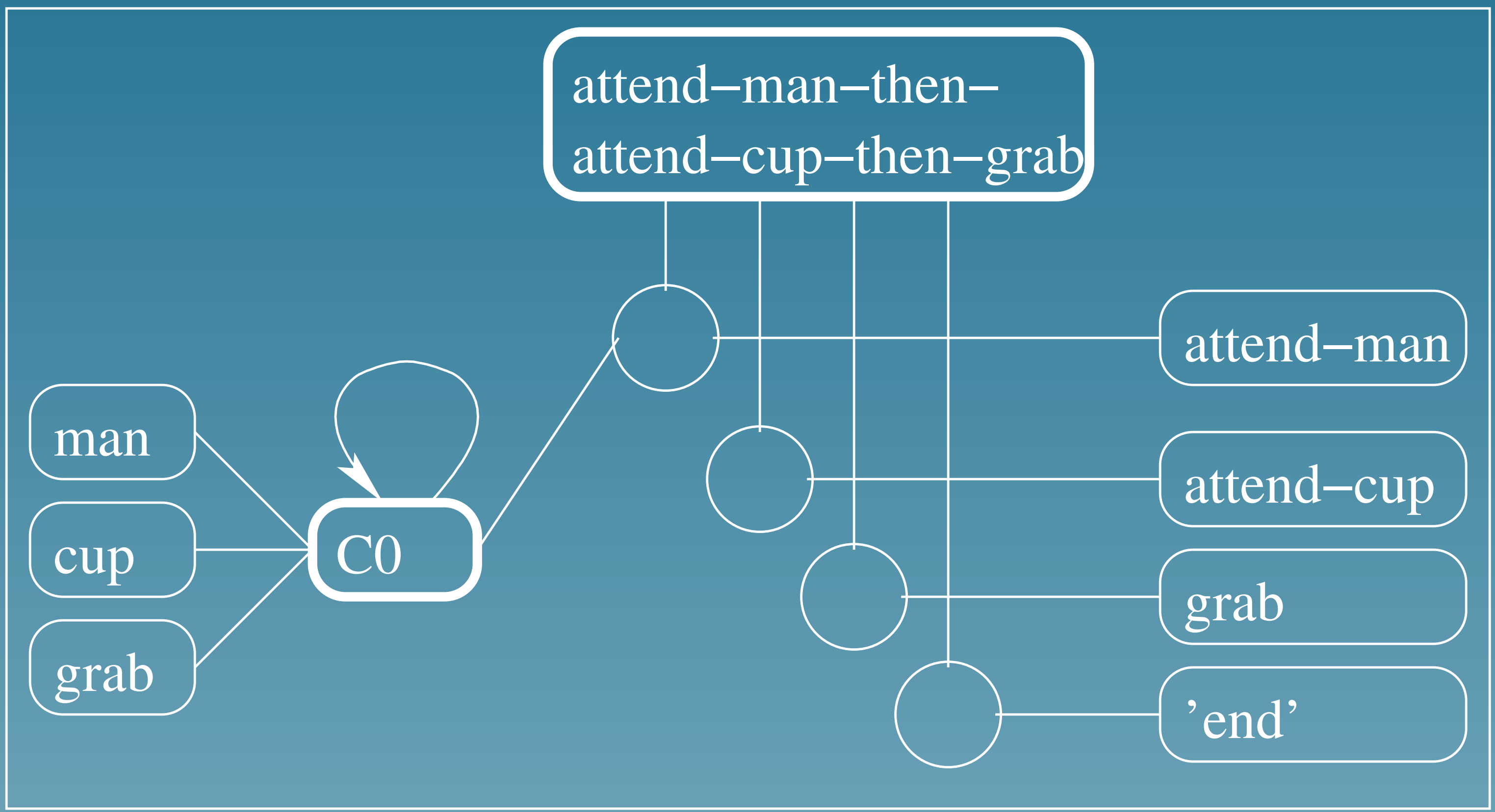
Suggestion: it's an active representation, consisting of the 'playing' of a SM sequence stored in working memory.

- There's a tonic component: the PFC sequence plan
- There's a phasic component: a sequence of SM states and contexts.

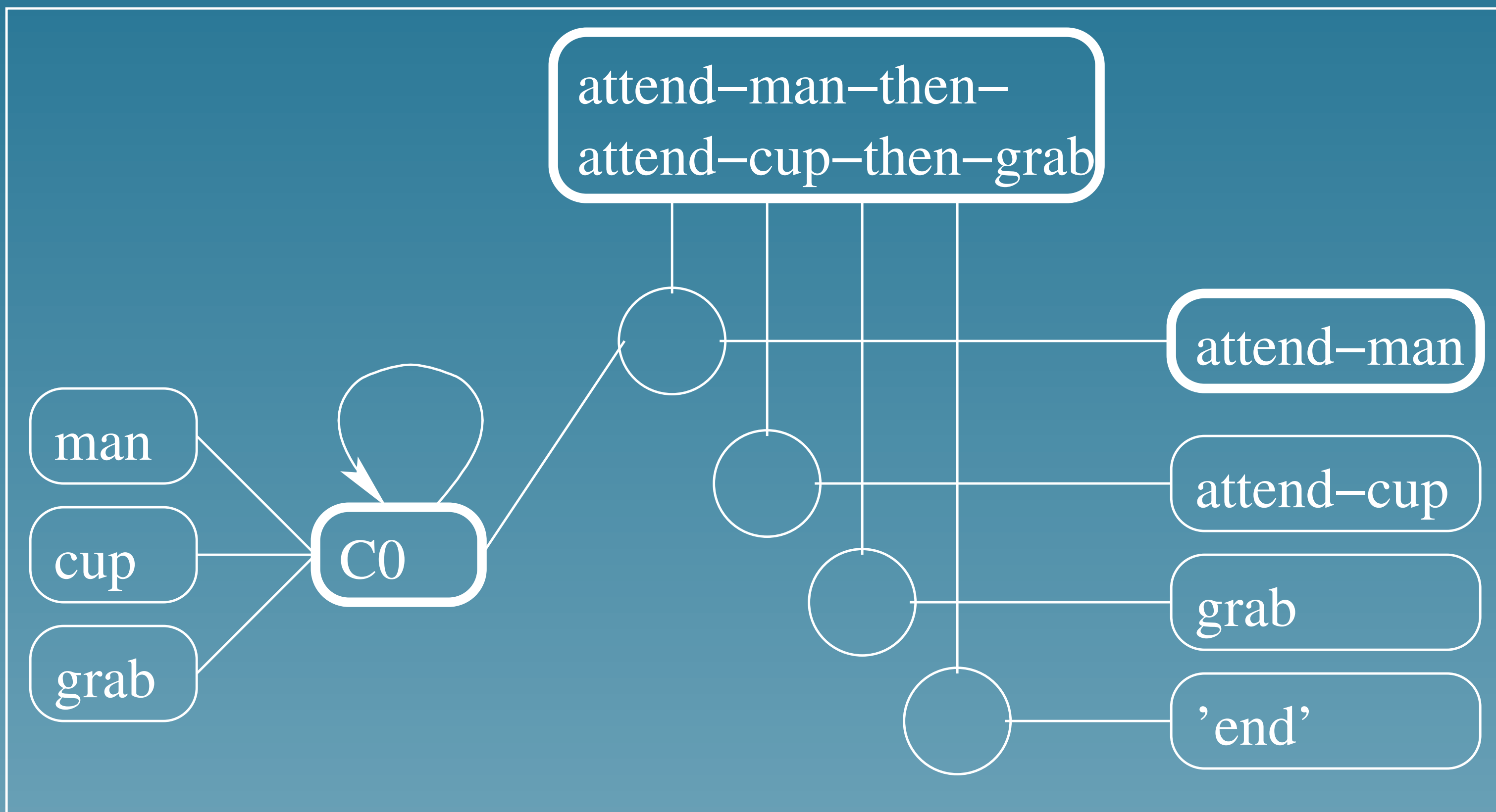
SM sequence for *The man grabbed the cup*



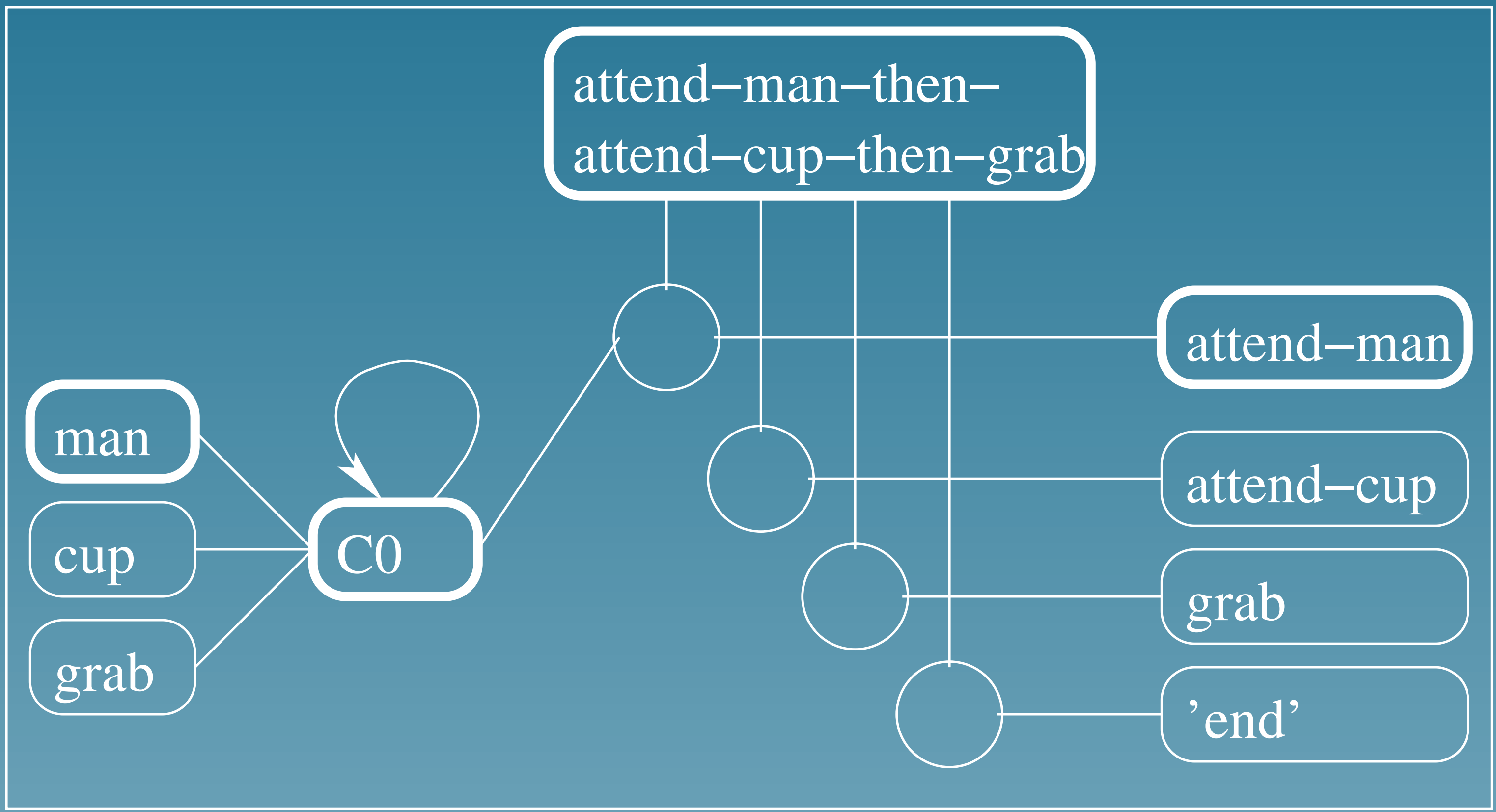
SM sequence for *The man grabbed the cup*



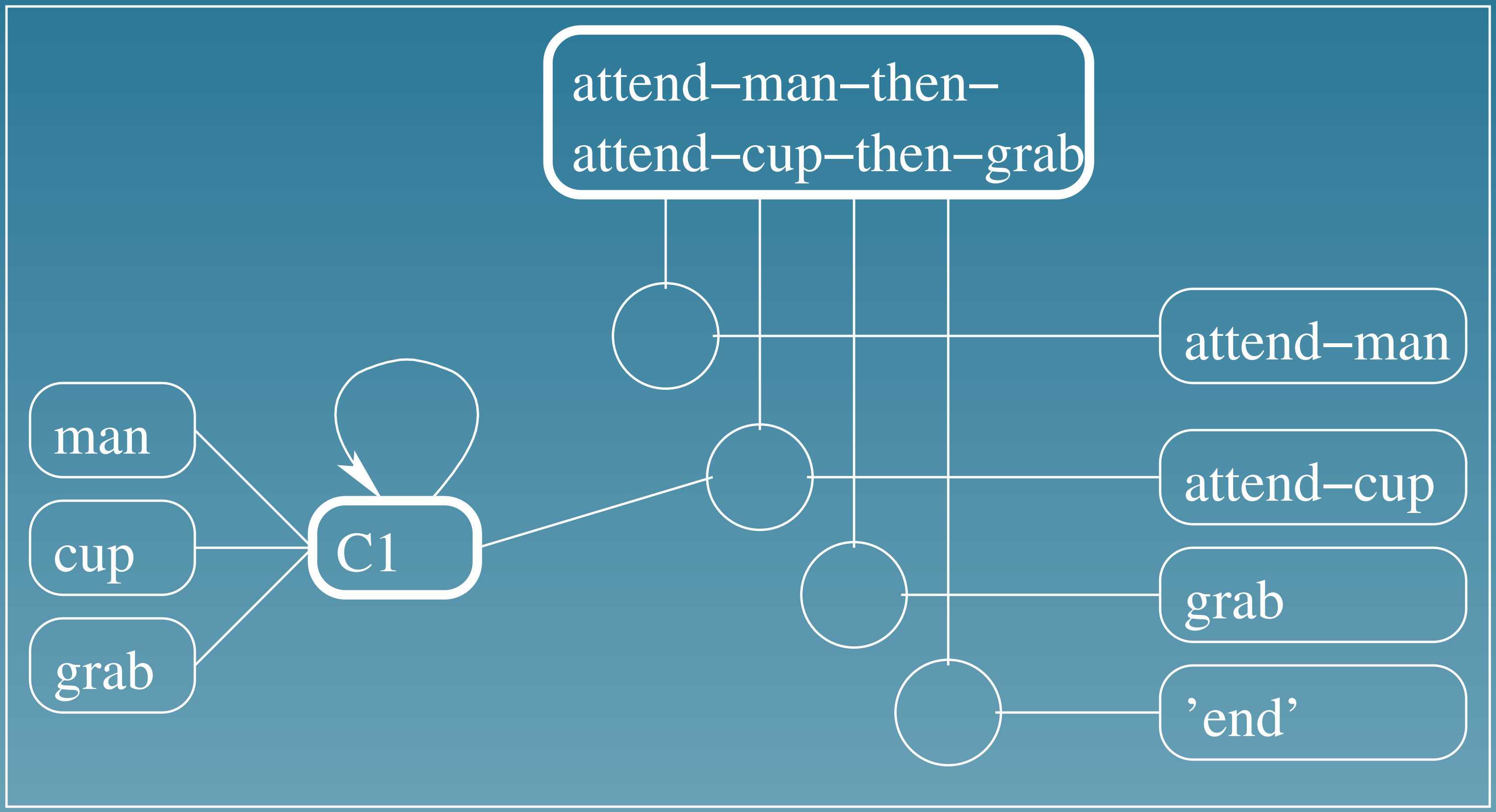
SM sequence for *The man grabbed the cup*



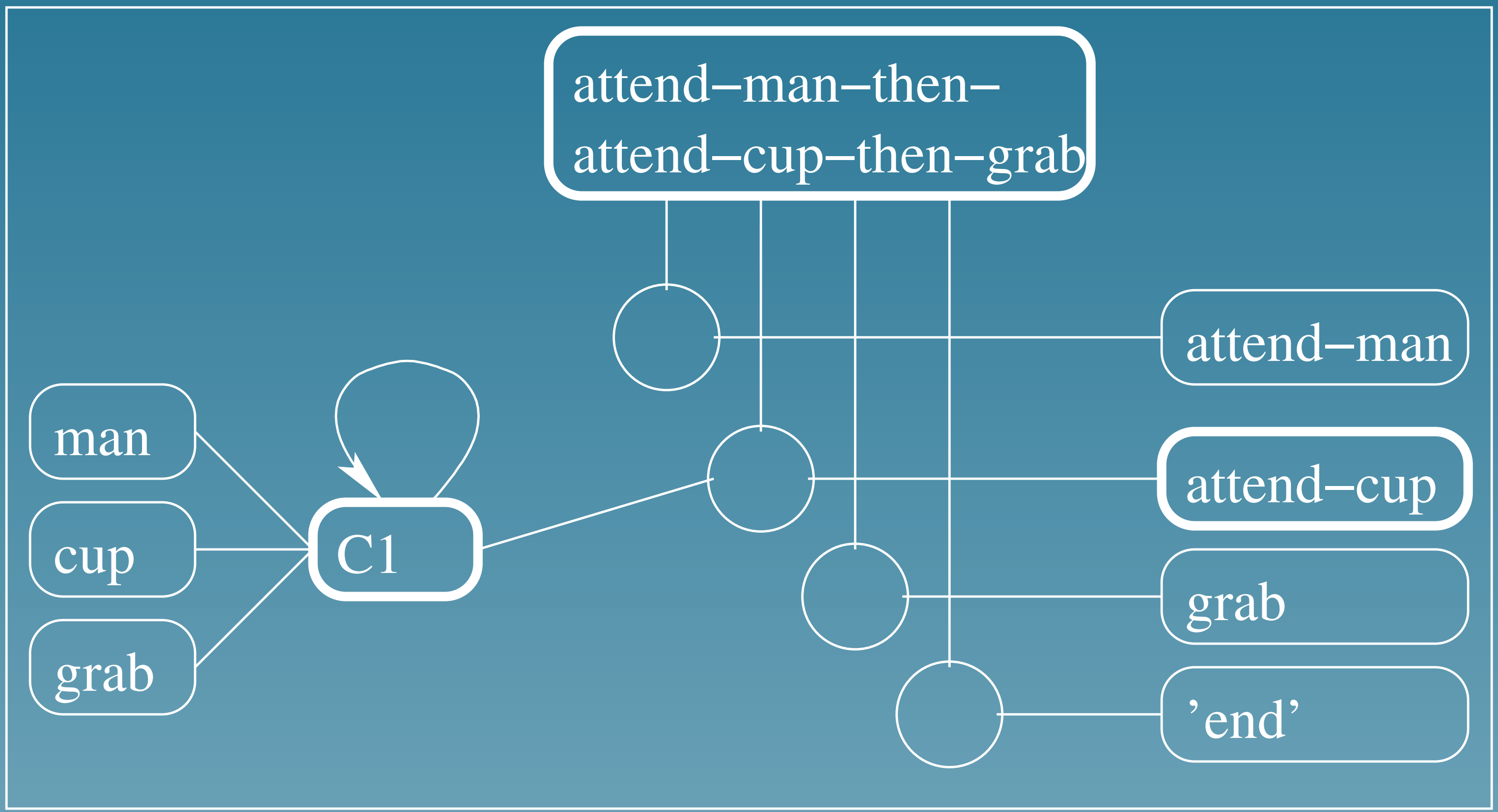
SM sequence for *The man grabbed the cup*



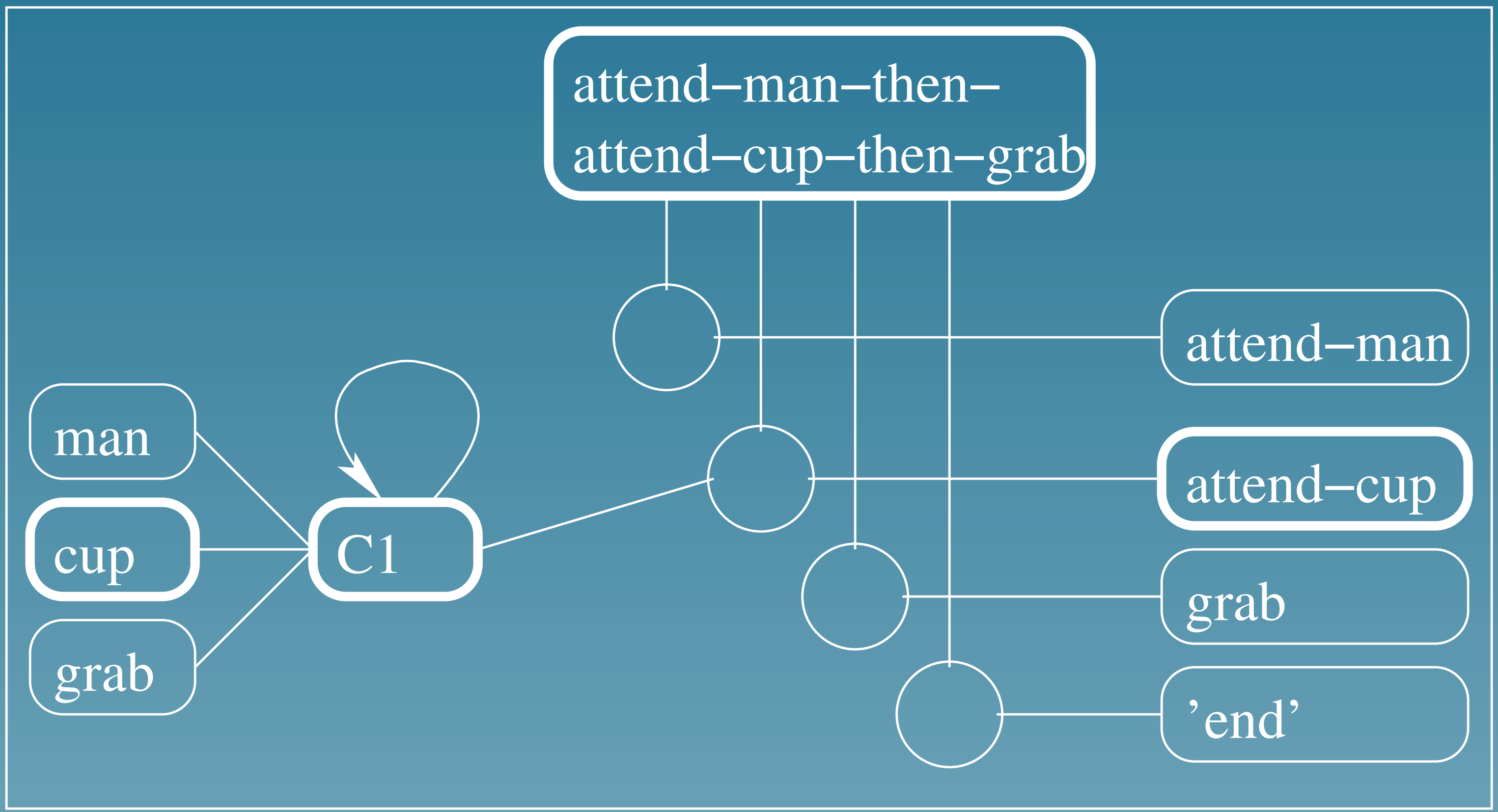
SM sequence for *The man grabbed the cup*



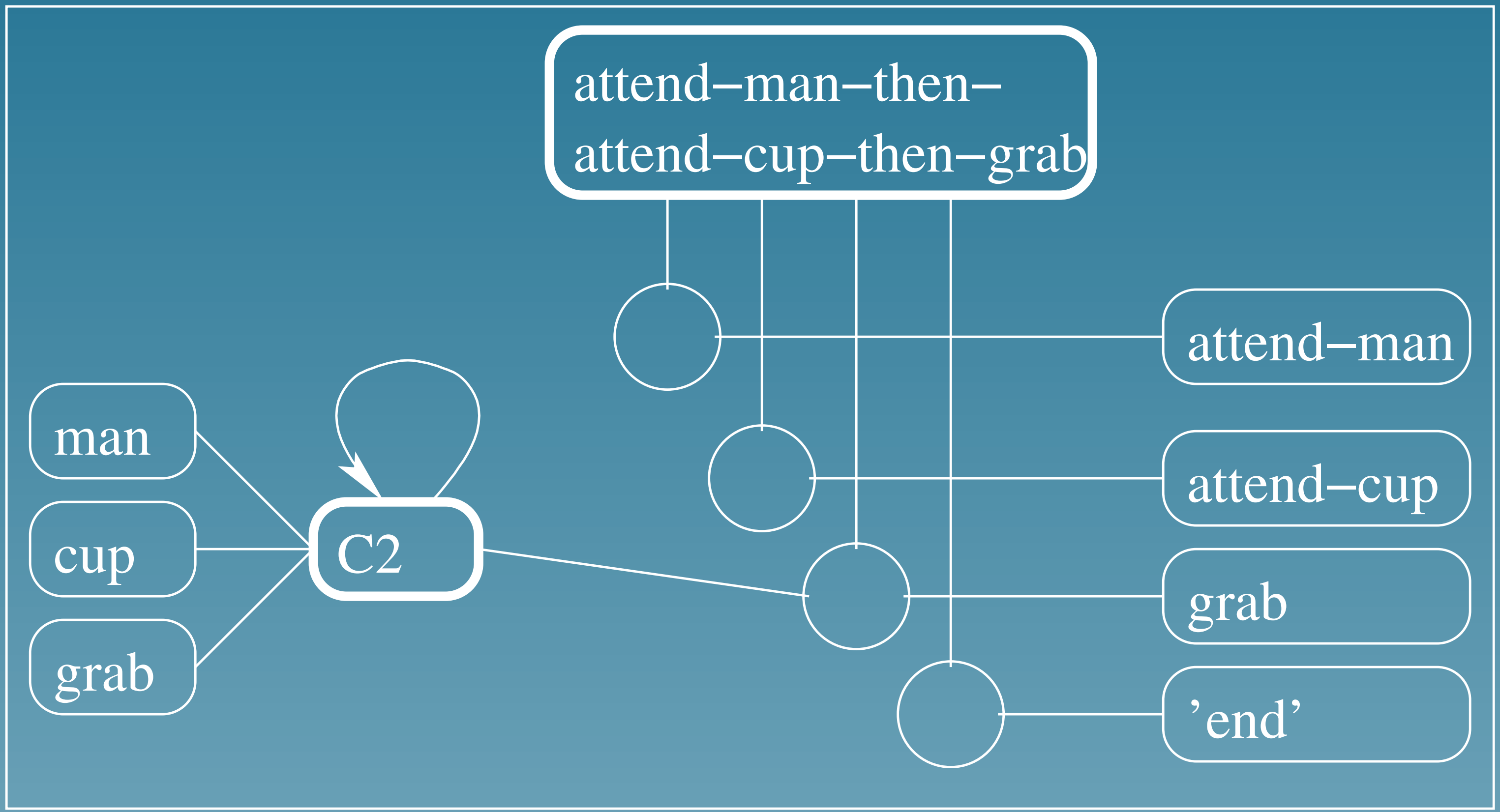
SM sequence for *The man grabbed the cup*



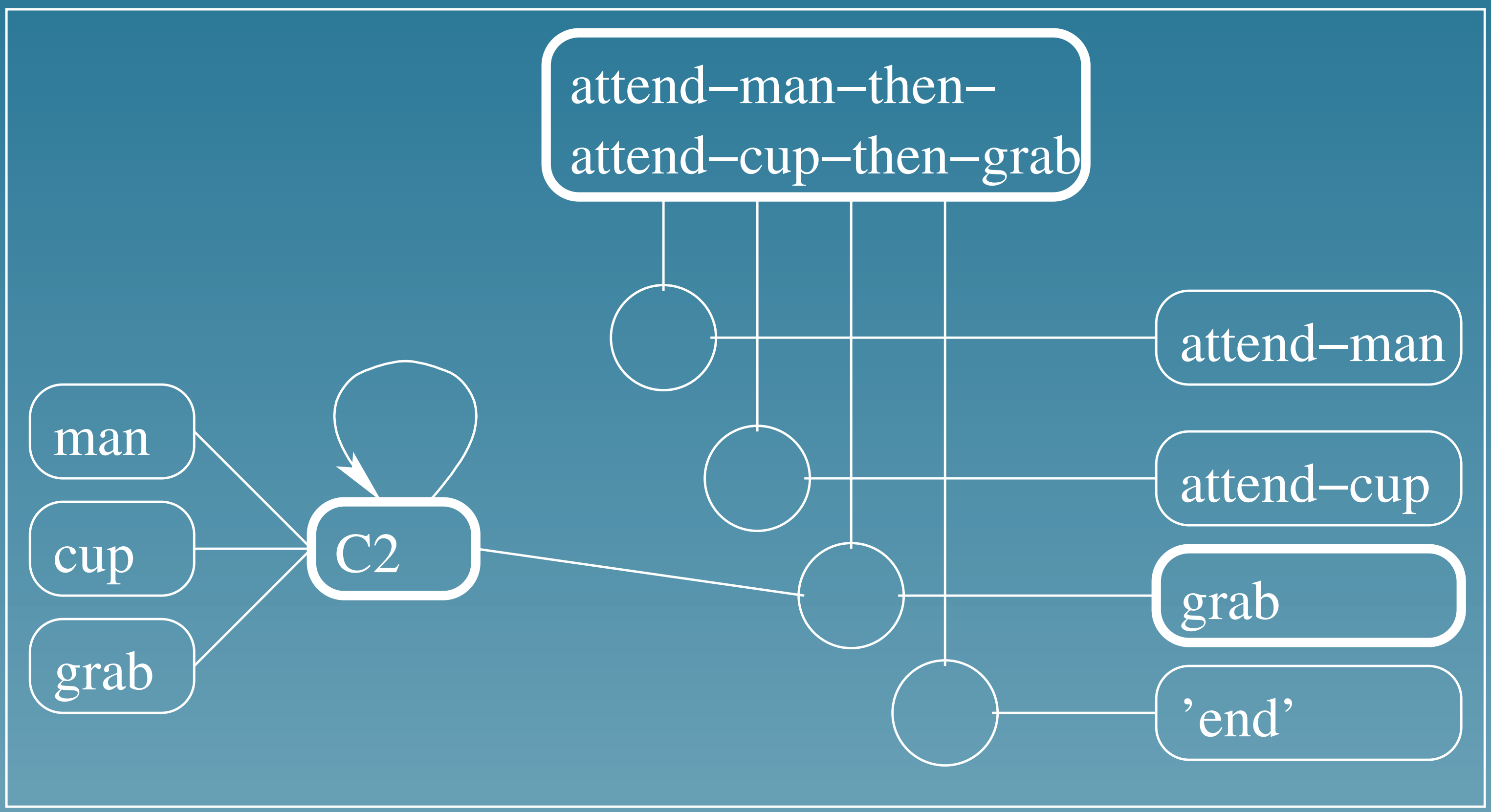
SM sequence for *The man grabbed the cup*



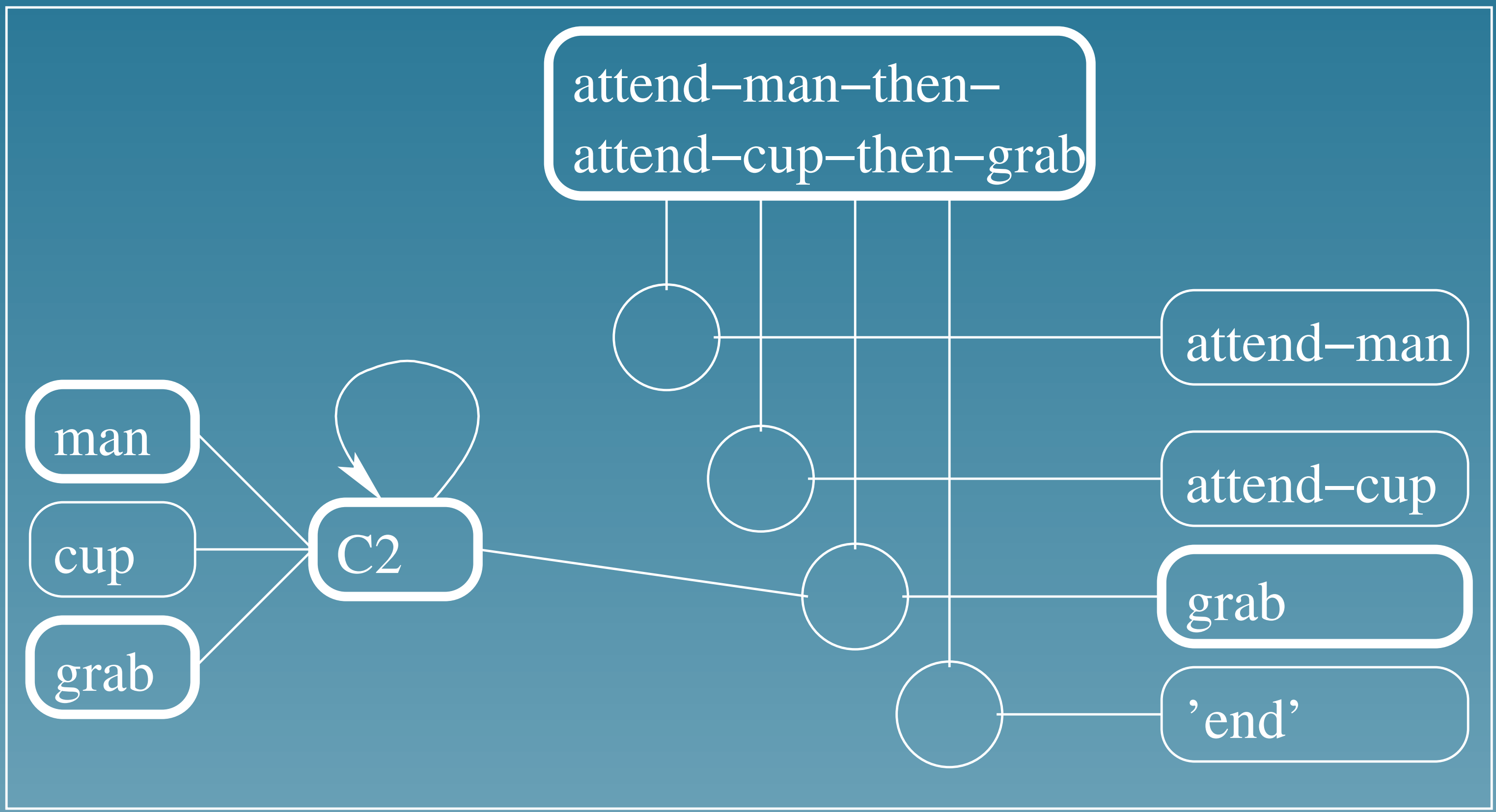
SM sequence for *The man grabbed the cup*



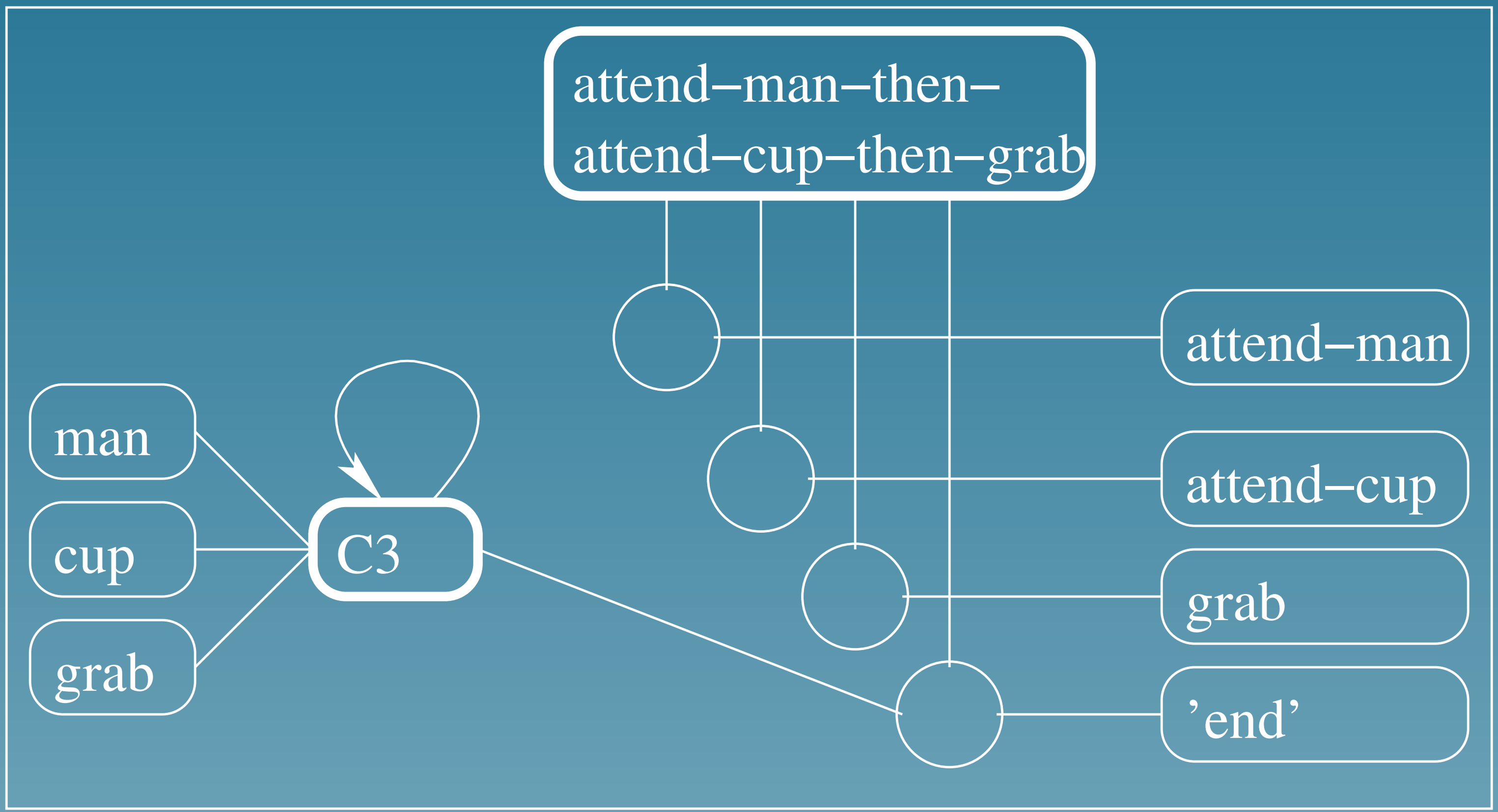
SM sequence for *The man grabbed the cup*



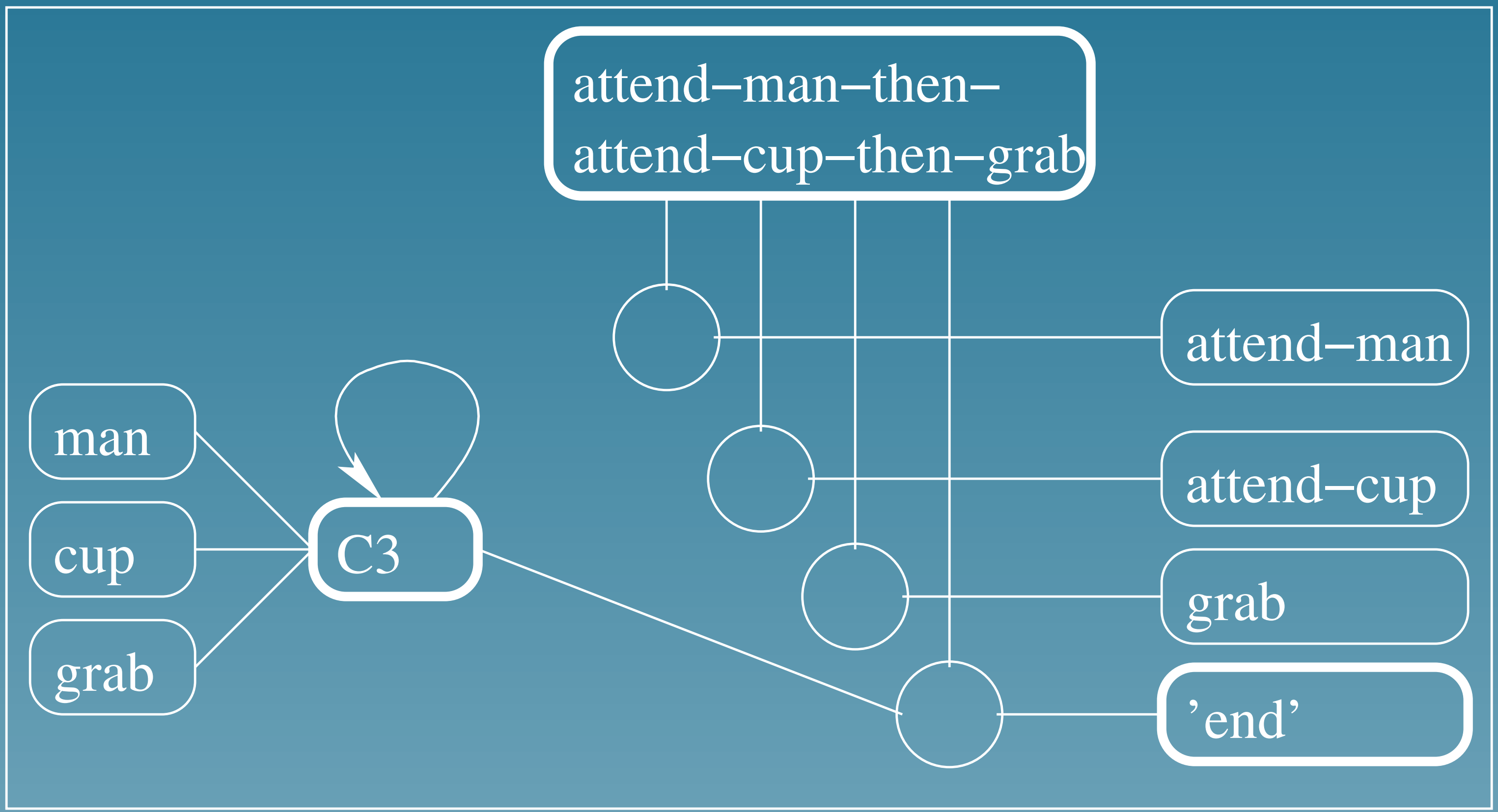
SM sequence for *The man grabbed the cup*



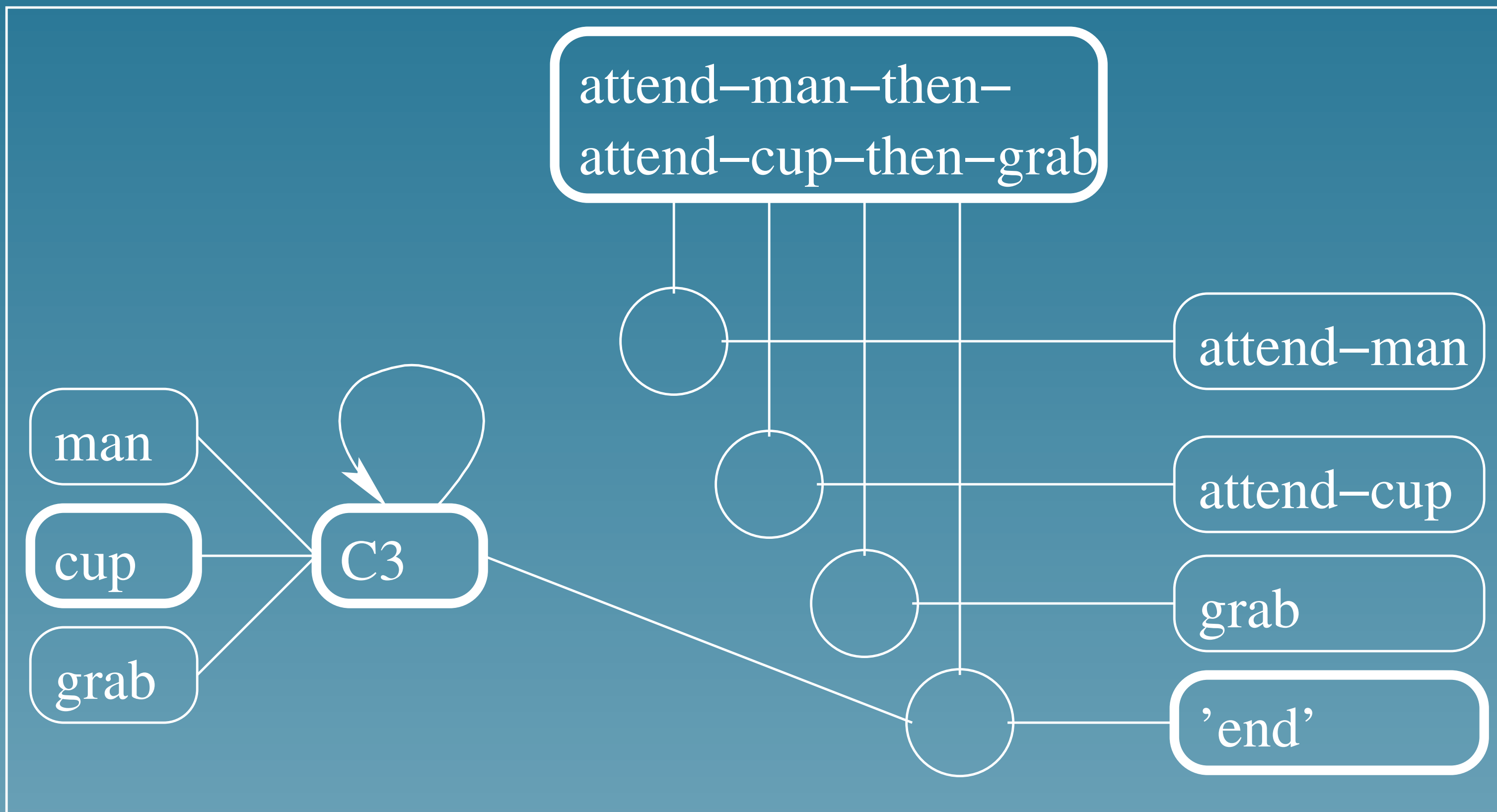
SM sequence for *The man grabbed the cup*



SM sequence for *The man grabbed the cup*



SM sequence for *The man grabbed the cup*



A model of sentence syntax

A model of sentence syntax

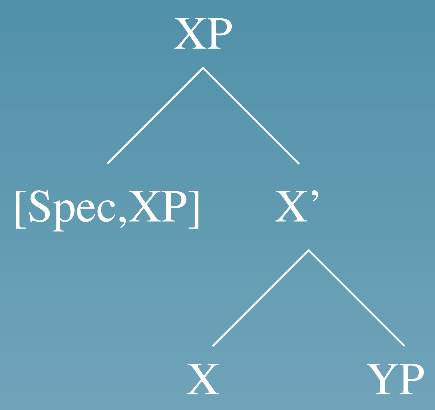
The syntactic framework I'm using is **Minimalism** (Chomsky, 1995). Very briefly:

- Sentences have a **phonetic form** (**PF**) and an underlying **logical form** (**LF**).

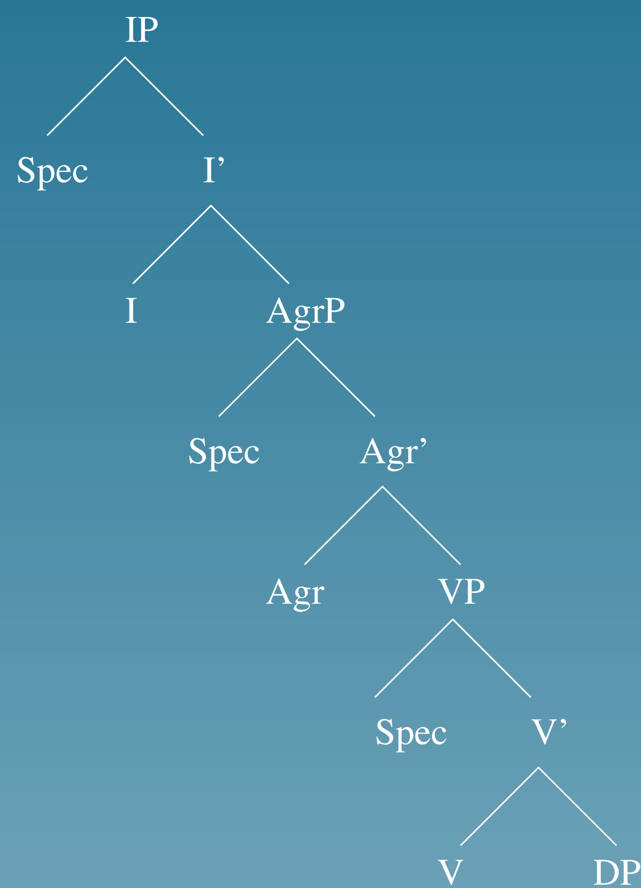
A model of sentence syntax

The syntactic framework I'm using is **Minimalism** (Chomsky, 1995). Very briefly:

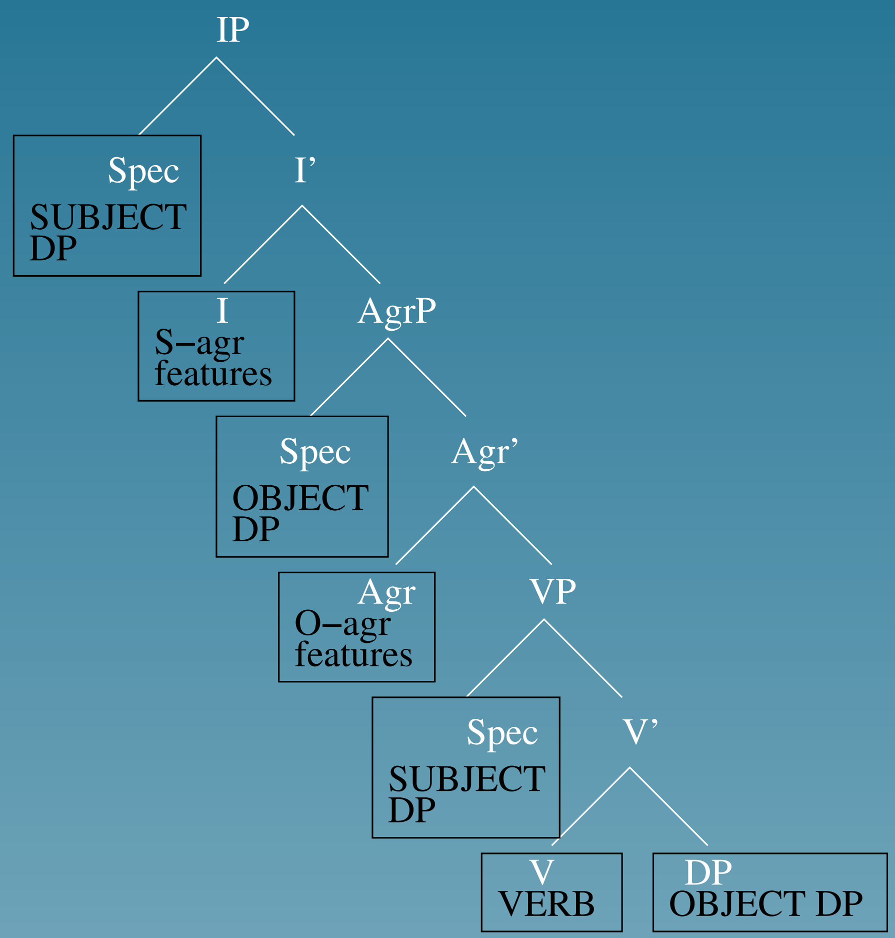
- Sentences have a **phonetic form** (**PF**) and an underlying **logical form** (**LF**).
- The building block for LF and PF is the **X-bar schema**.



The structure of a transitive clause



The structure of a transitive clause



DP movement

In Minimalism, subject and object DPs need two things:

- a **thematic role** (e.g. AGENT or PATIENT)
- **Case** (e.g. NOM or ACC).

They get their thematic role within the VP.

They need to *move* to higher Spec positions to get Case. (They're *assigned* Case by I and Agr.)

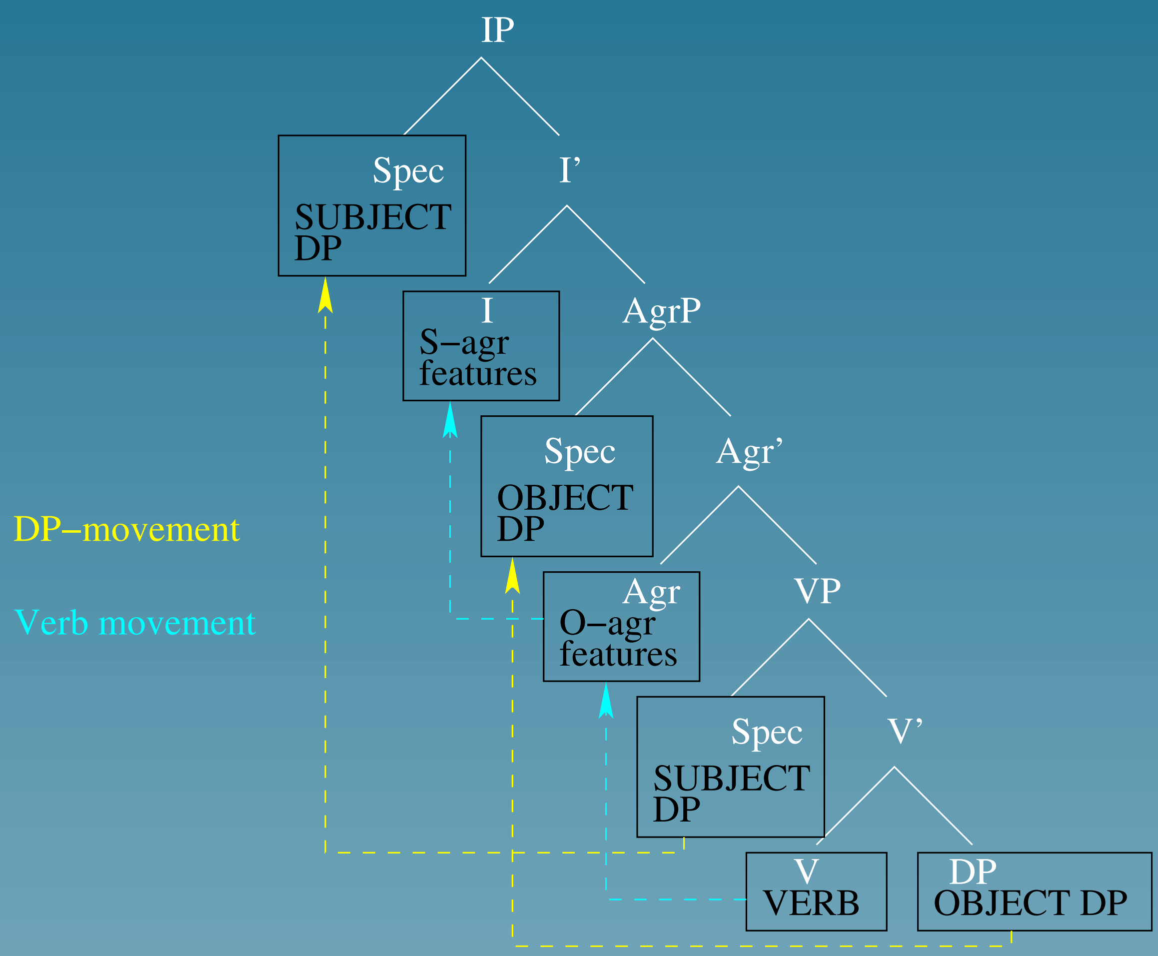
V movement

The verb originates at V.
Its *inflections* originate at Agr and I.

- *In finite sentences*, V is inflected:
  $The\ man\ grabbed\ a\ cup$
  V *moves* to Agr and then I to get its inflections.

- *In nonfinite sentences*, V is uninflected:
  $The\ man\ wants\ to\ /\ tries\ to\ /\ can\ [grab\ a\ cup]$
  V doesn't move.

Movement at LF for a transitive clause



The mapping from LF to PF

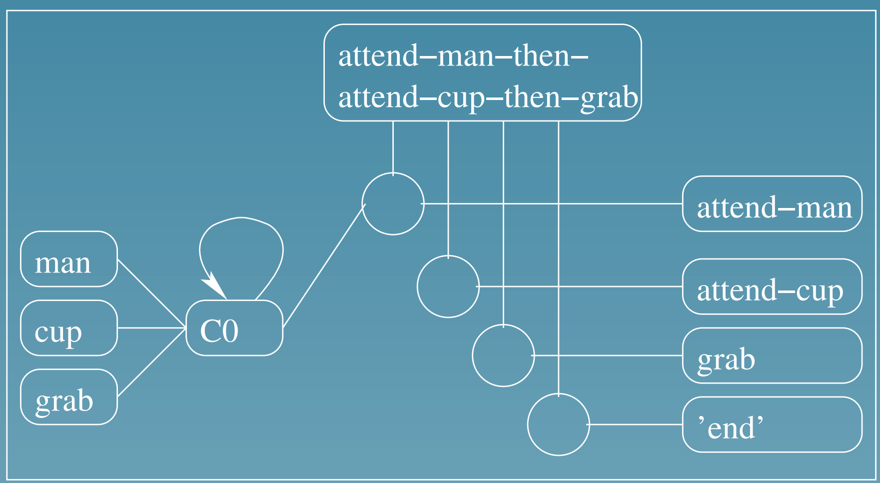
At some point during these movements, the PF of the sentence is ‘read off’ the LF tree.

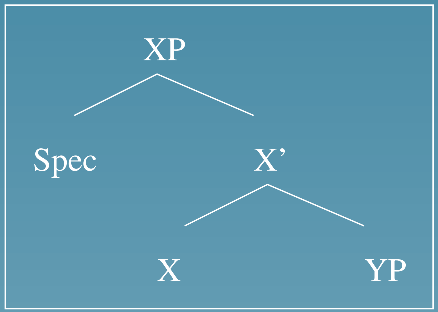
Different orderings of words in different languages reflect the time at which PF is read off from LF.

LF	Spec	I	Spec	Agr	Spec	V	DP
English PF	man					grab	cup
French PF	man	grab					cup
Maori PF		grab			man		cup

A sensorimotor interpretation of LF

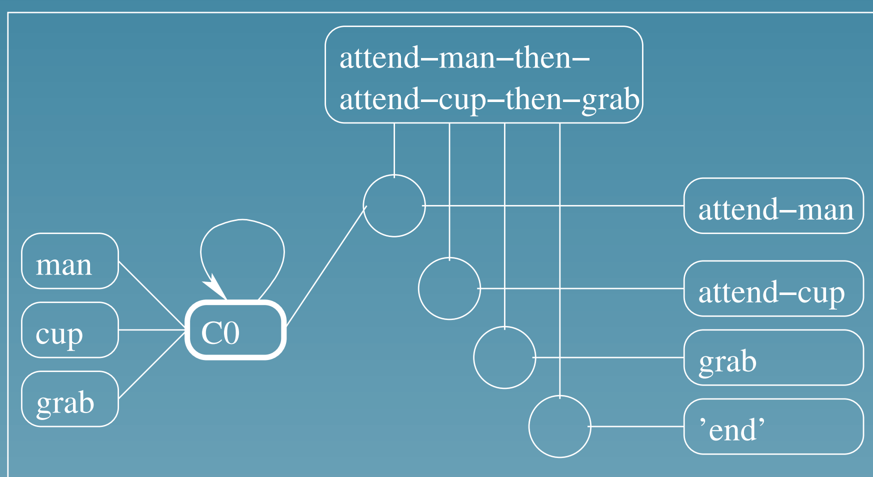
Main idea: the right-branching X-bar structure of LF is an encoding of the representations in successive cycles of the sensorimotor network.

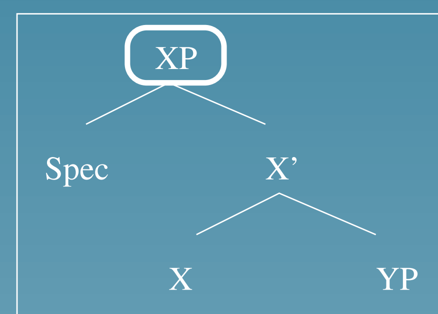




A sensorimotor interpretation of LF

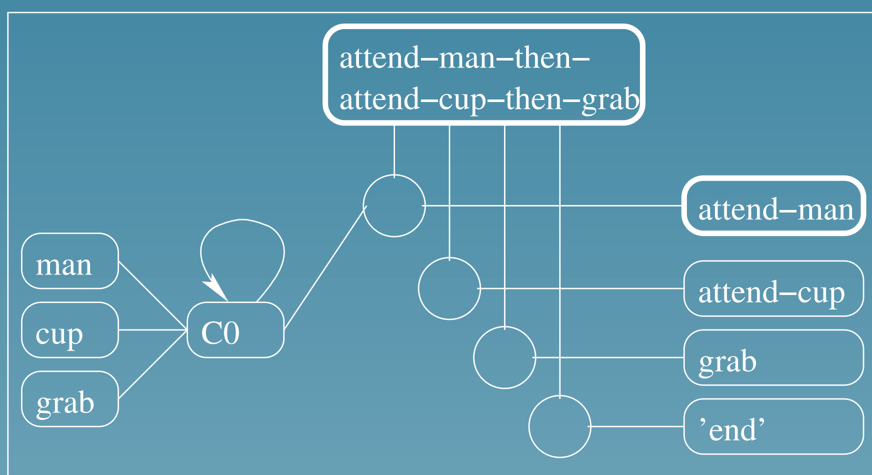
Main idea: the right-branching X-bar structure of LF is an encoding of the representations in successive cycles of the sensorimotor network.

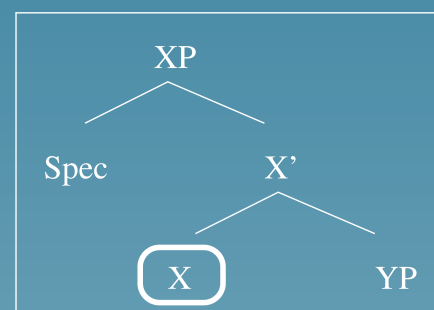




A sensorimotor interpretation of LF

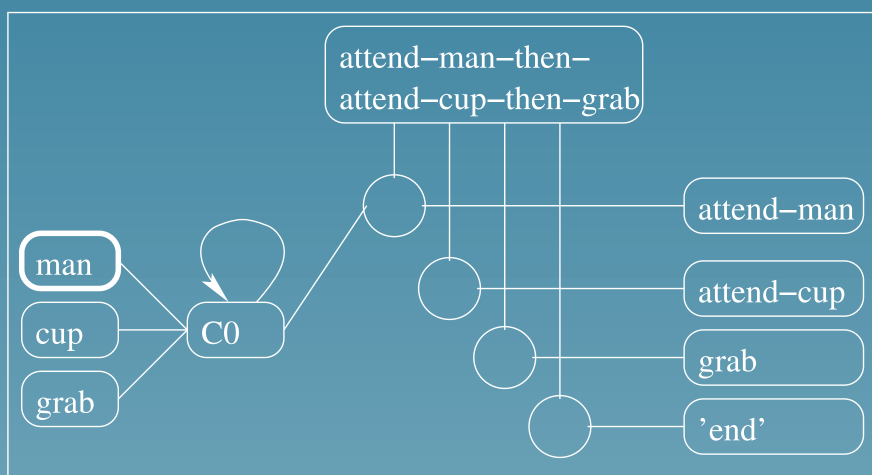
Main idea: the right-branching X-bar structure of LF is an encoding of the representations in successive cycles of the sensorimotor network.

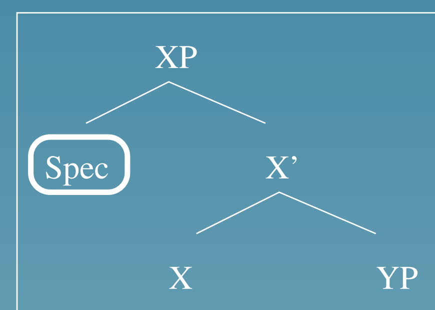




A sensorimotor interpretation of LF

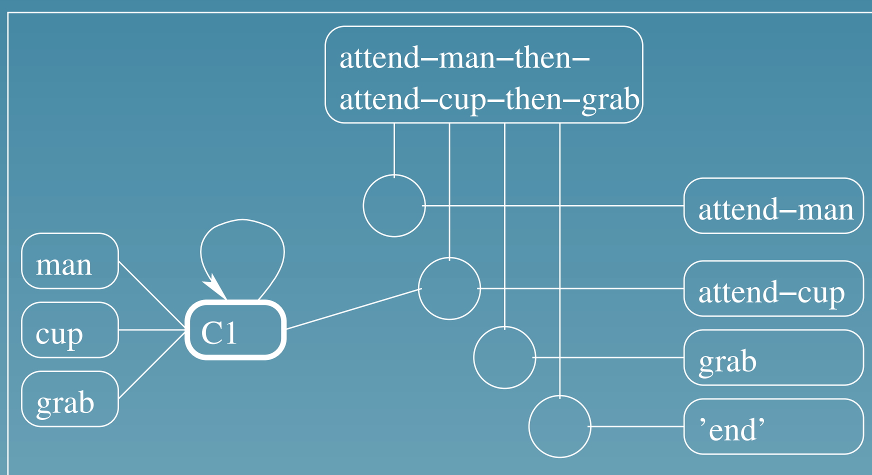
Main idea: the right-branching X-bar structure of LF is an encoding of the representations in successive cycles of the sensorimotor network.

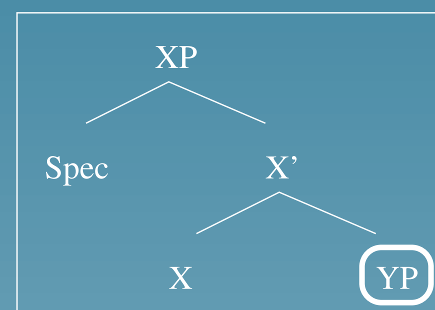




A sensorimotor interpretation of LF

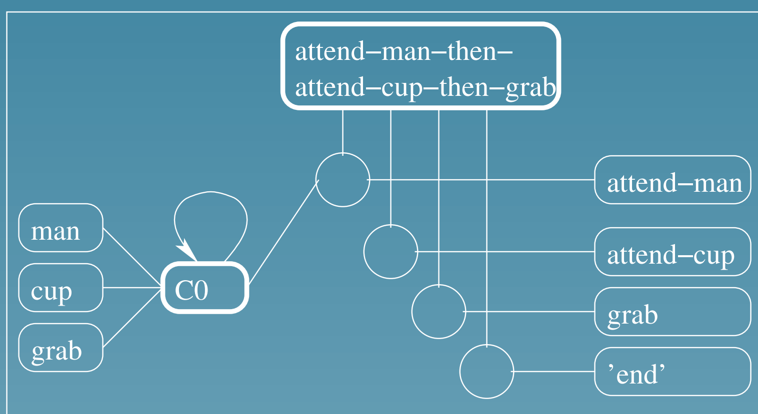
Main idea: the right-branching X-bar structure of LF is an encoding of the representations in successive cycles of the sensorimotor network.

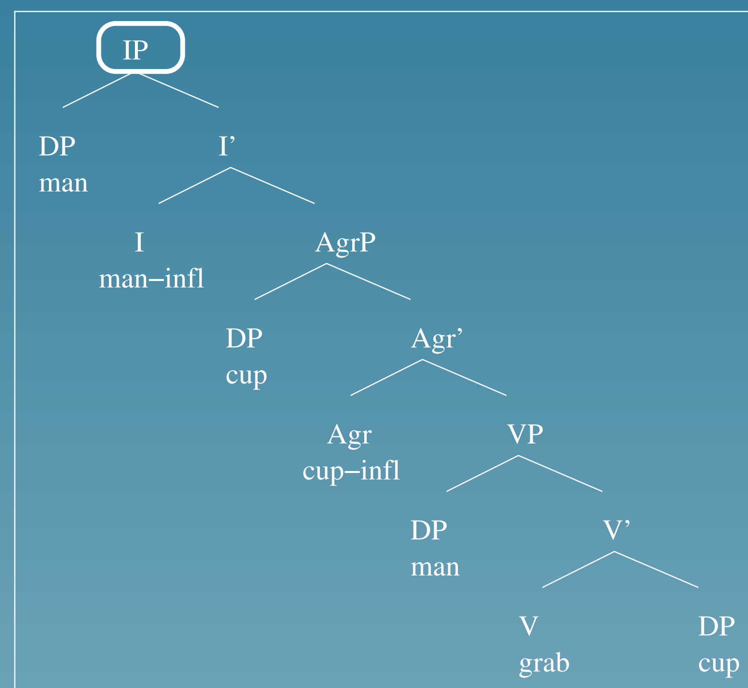




A sensorimotor interpretation of LF

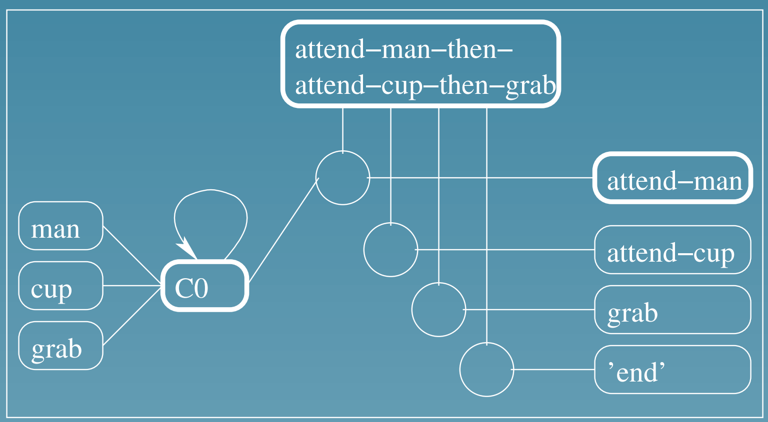
The LF of *The man grabbed the cup* is understood as a rehearsed SM sequence, as follows:

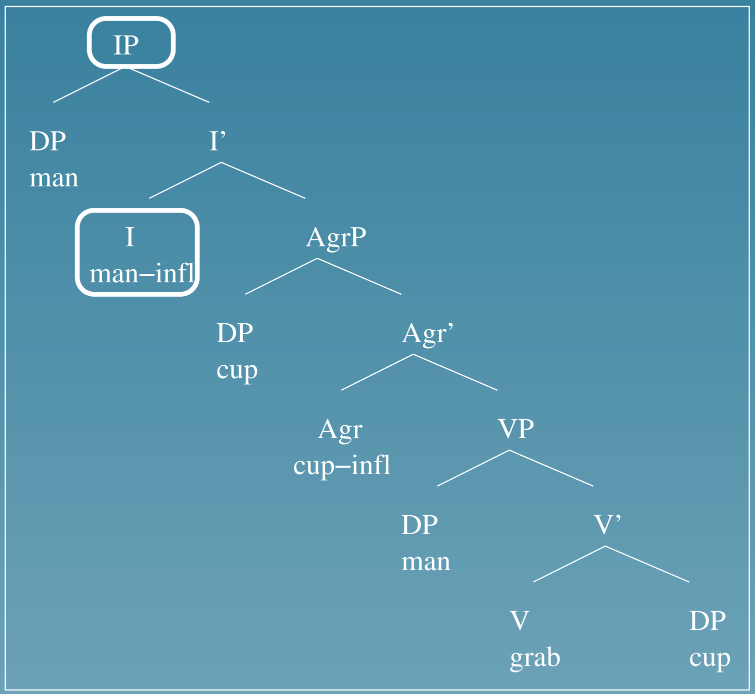




A sensorimotor interpretation of LF

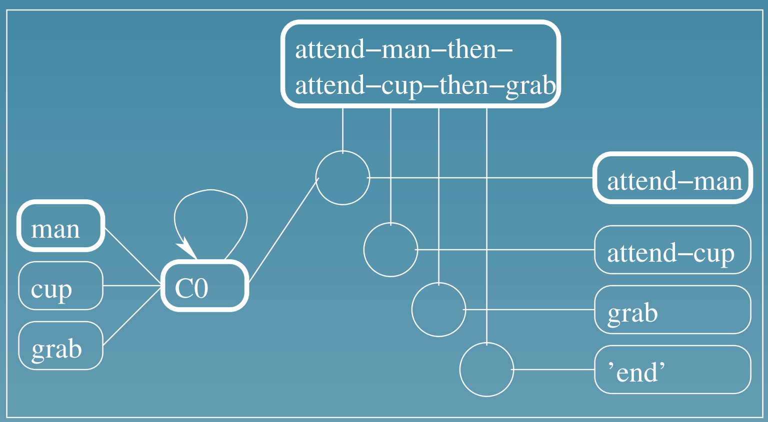
The LF of *The man grabbed the cup* is understood as a rehearsed SM sequence, as follows:

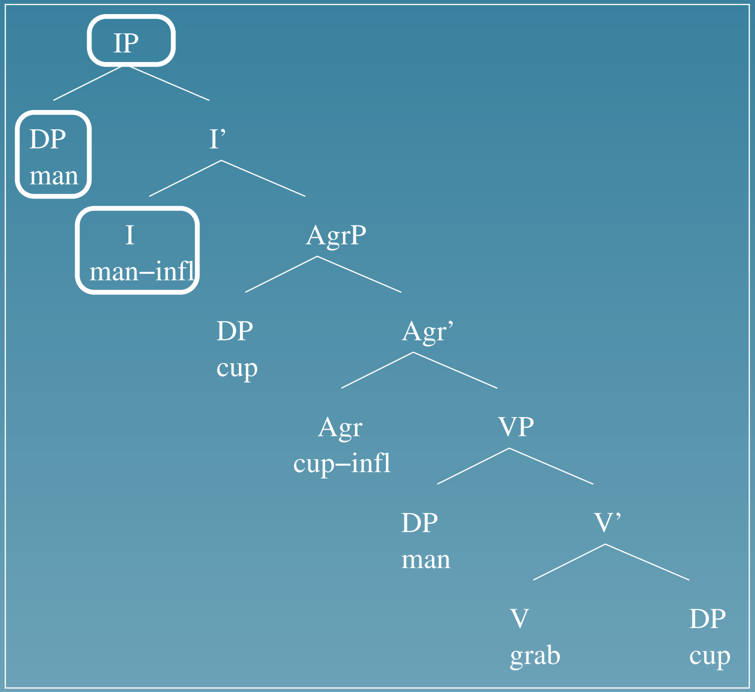




A sensorimotor interpretation of LF

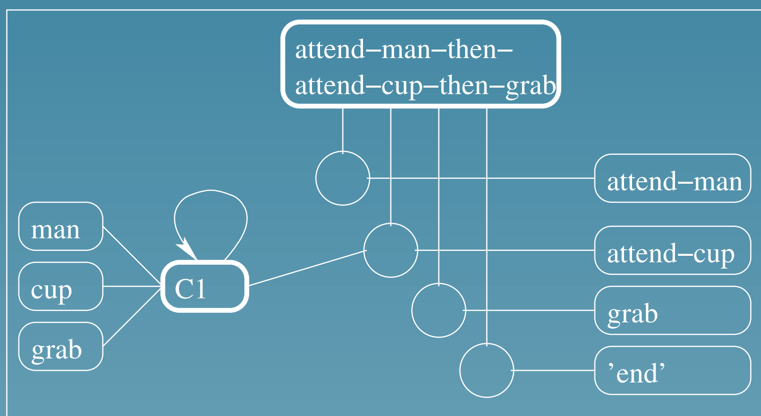
The LF of *The man grabbed the cup* is understood as a rehearsed SM sequence, as follows:

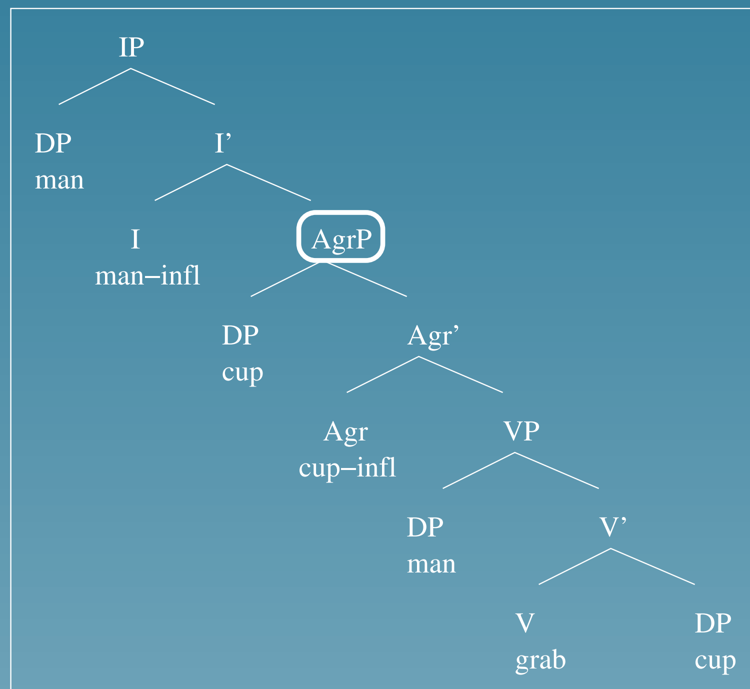




A sensorimotor interpretation of LF

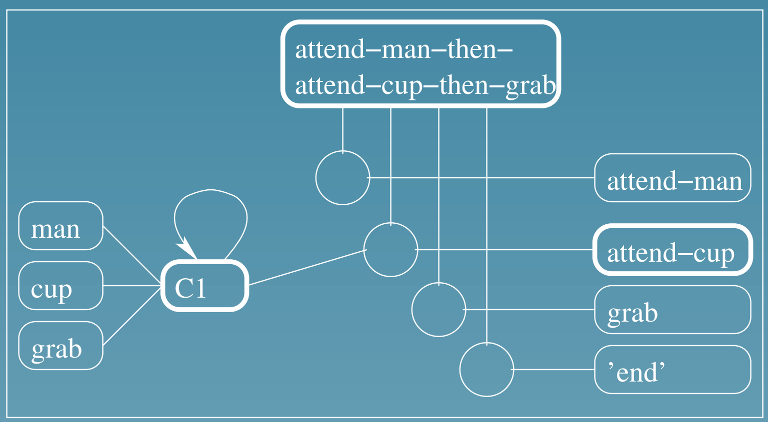
The LF of *The man grabbed the cup* is understood as a rehearsed SM sequence, as follows:

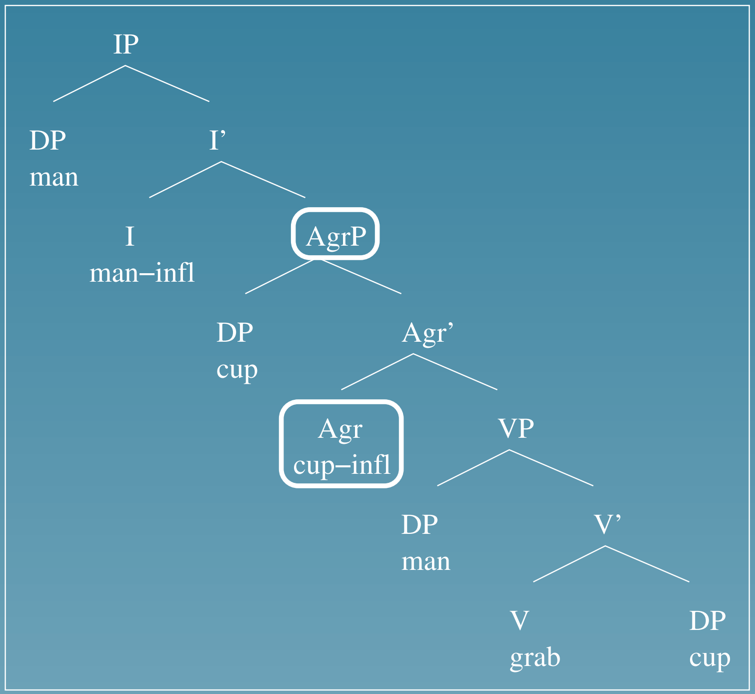




A sensorimotor interpretation of LF

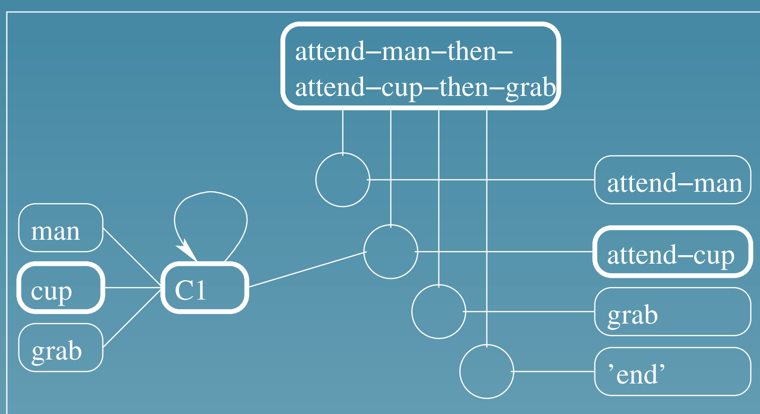
The LF of *The man grabbed the cup* is understood as a rehearsed SM sequence, as follows:

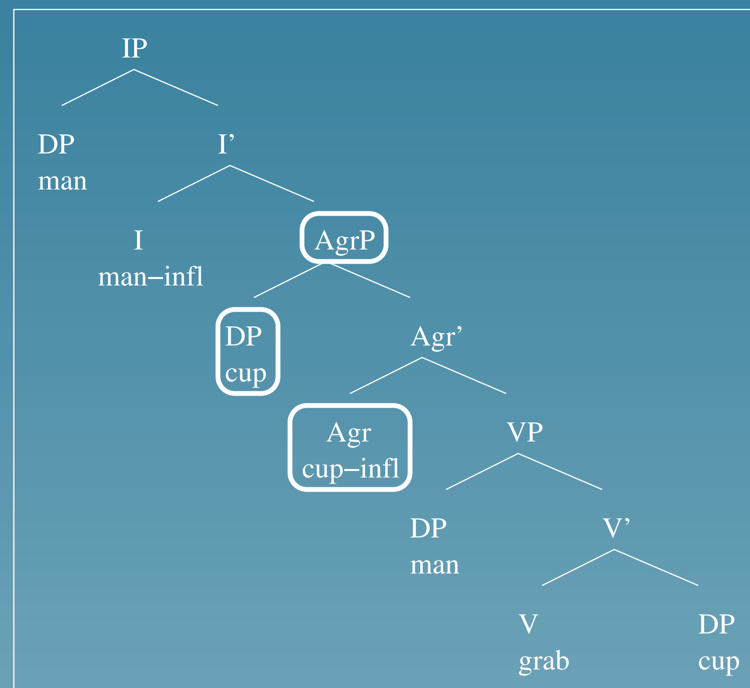




A sensorimotor interpretation of LF

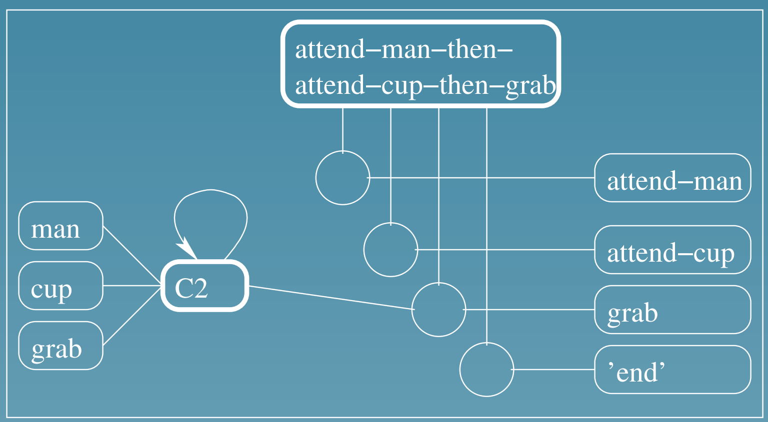
The LF of *The man grabbed the cup* is understood as a rehearsed SM sequence, as follows:

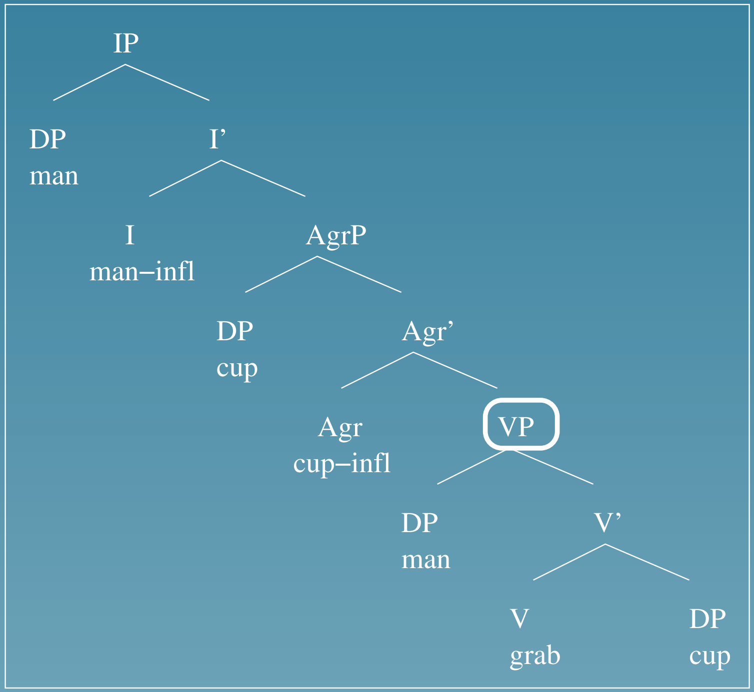




A sensorimotor interpretation of LF

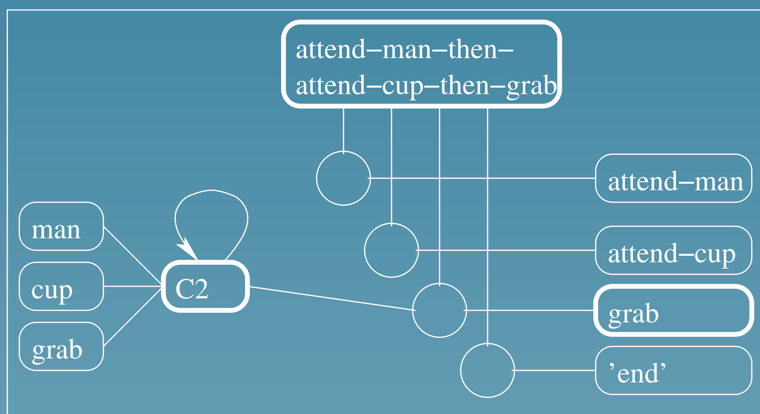
The LF of *The man grabbed the cup* is understood as a rehearsed SM sequence, as follows:

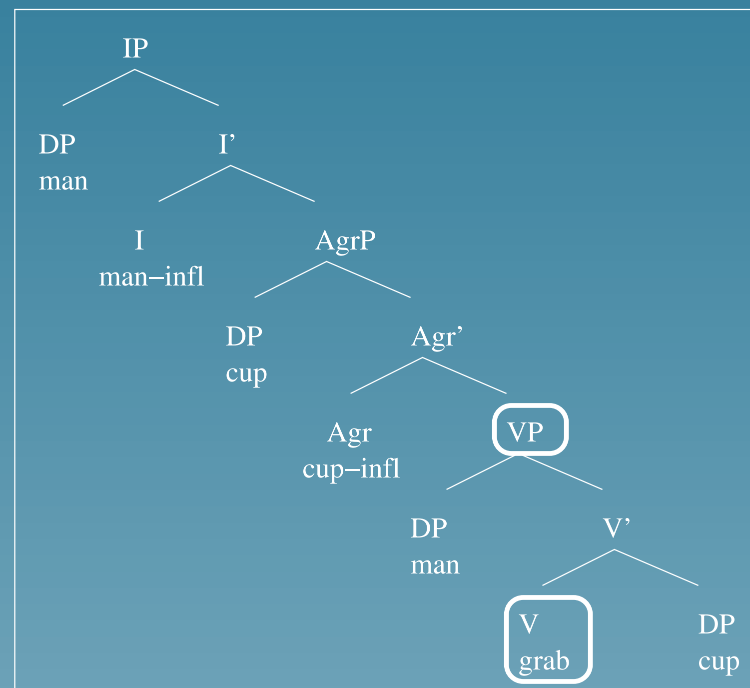




A sensorimotor interpretation of LF

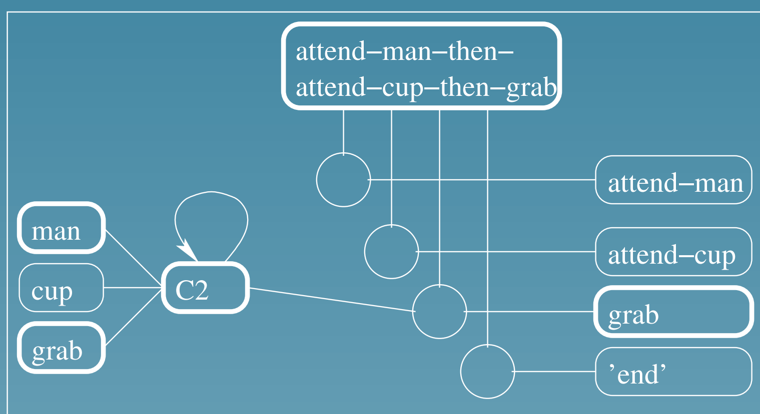
The LF of *The man grabbed the cup* is understood as a rehearsed SM sequence, as follows:

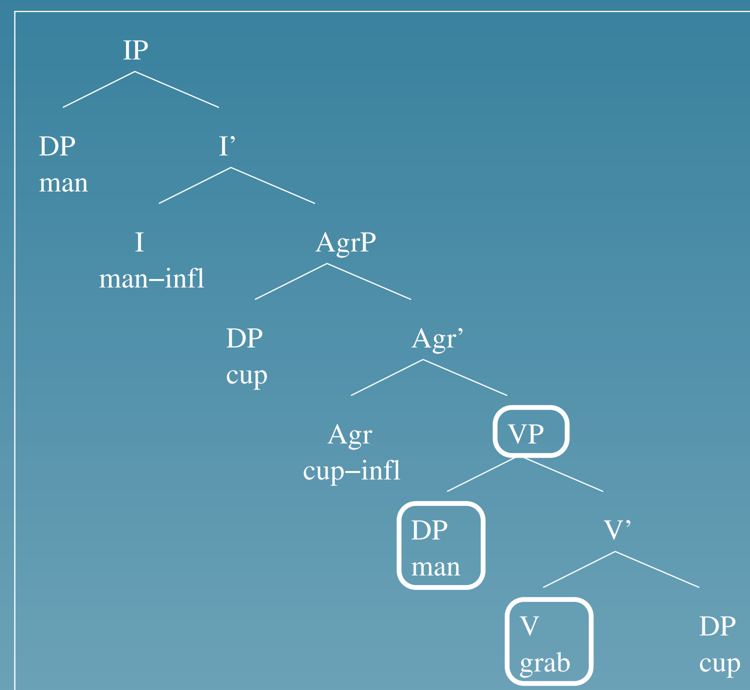




A sensorimotor interpretation of LF

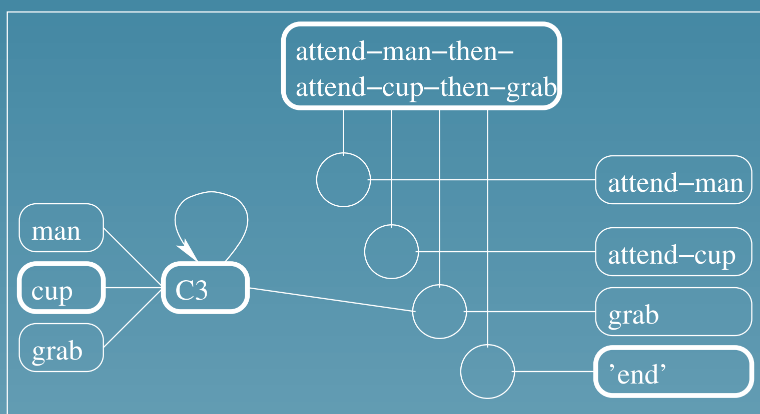
The LF of *The man grabbed the cup* is understood as a rehearsed SM sequence, as follows:

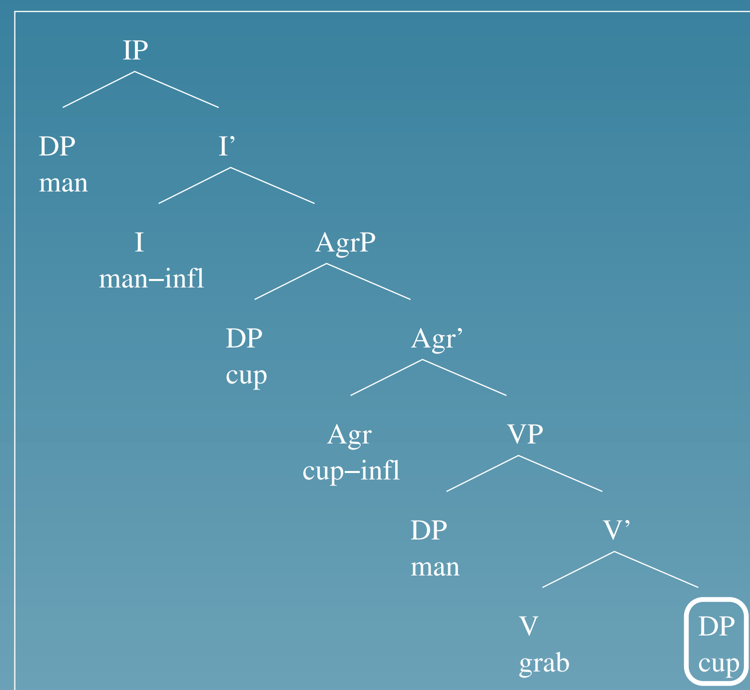




A sensorimotor interpretation of LF

The LF of *The man grabbed the cup* is understood as a rehearsed SM sequence, as follows:





Overlaps between syntactic & SM models

Everything in the LF structure of a transitive clause can be given a natural sensorimotor interpretation.

Right-branching X-bar structure	Successive cycles of the SM network
Individual X-bar components	Individual representations in the SM network
DP-movement	Re-attention to agent and patient
V-movement	Tonic activation of actions in PFC sequence plan

Mapping from LF to PF

Assume: everyone has the same SM sequence for *The man grabbed a cup*.

- But different languages express the proposition using different word orderings.

The mapping from SM to word sequences must be *learned*. The training data is a set of pairs:

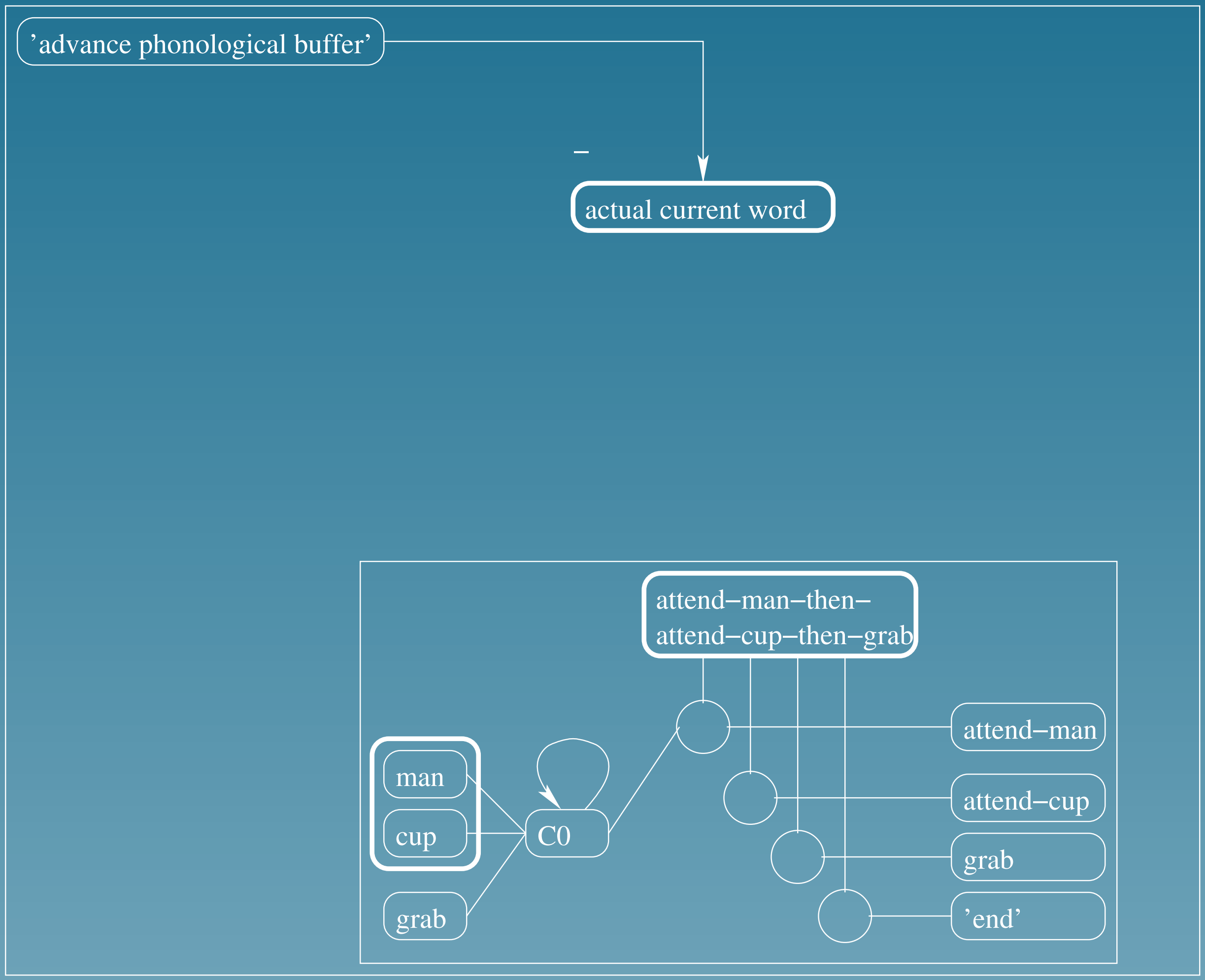
- SM sequence (held in episodic buffer)
- Phonological sequence (held in phonological buffer)

An LF-to-PF mapping network

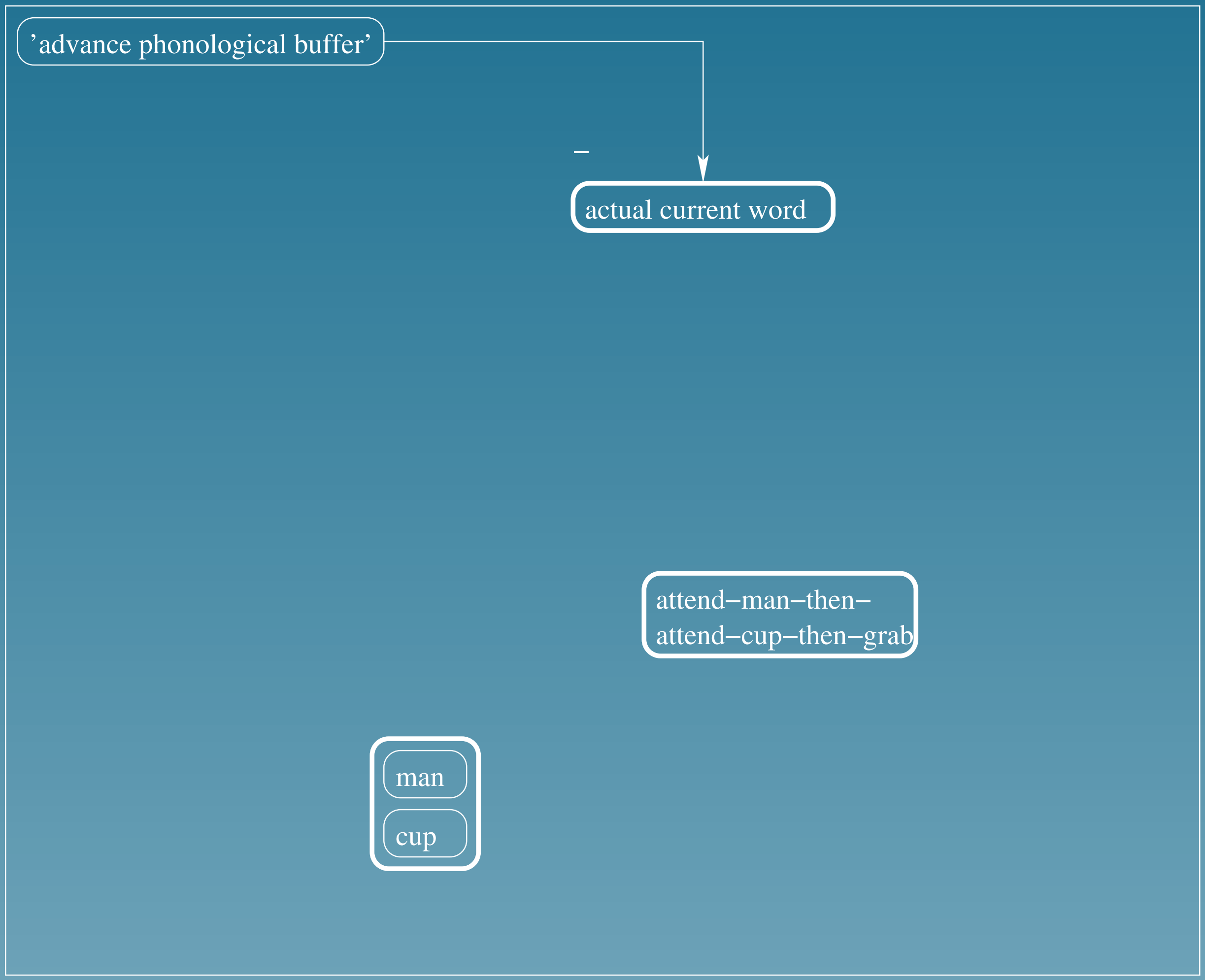
We have built a network which takes training pairs of this sort, and learns word ordering conventions.

- The network's task is to *predict the next word* in the input word sequence. (C.f. Elman, Dell, Bock, Chang. . .)
- It achieves this by learning when to generate 'gaps' (phonologically empty lexical items).
- The network maintains its own 'context' representation, to store the history of words in the sequence.

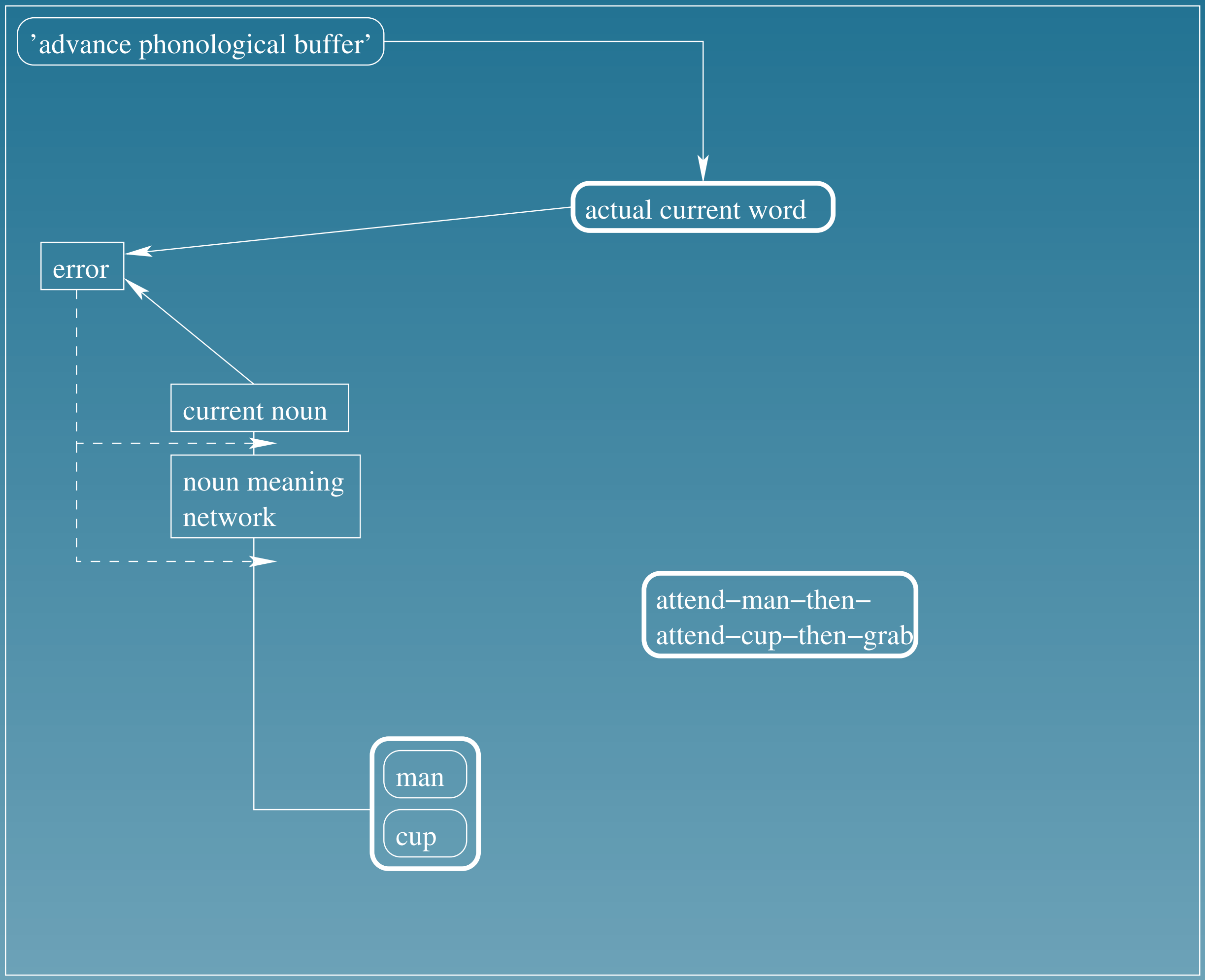
LF-to-PF mapping network: inputs



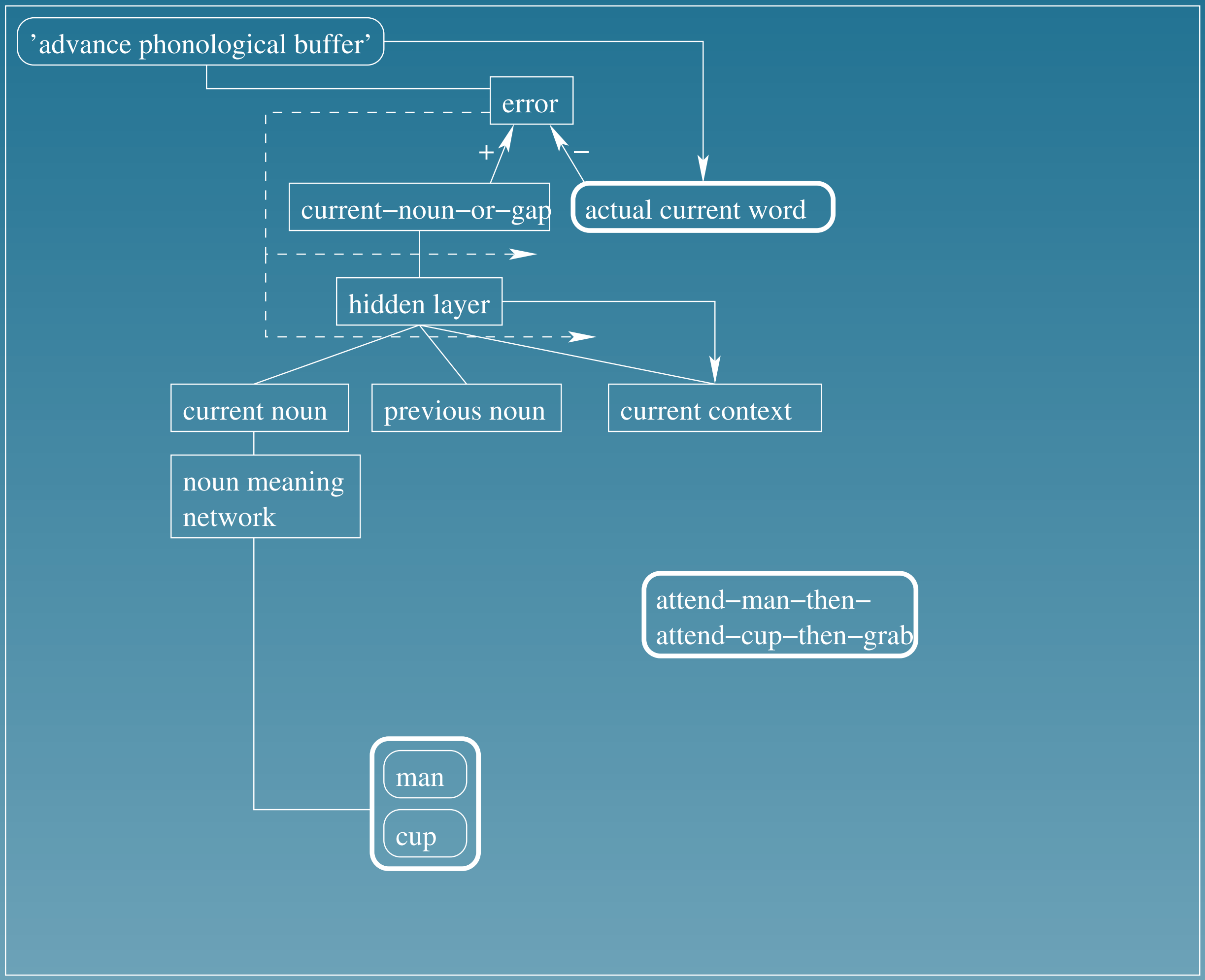
LF-to-PF mapping network: inputs



Noun semantics network



Noun sequencing network



Results of the noun sequencing network

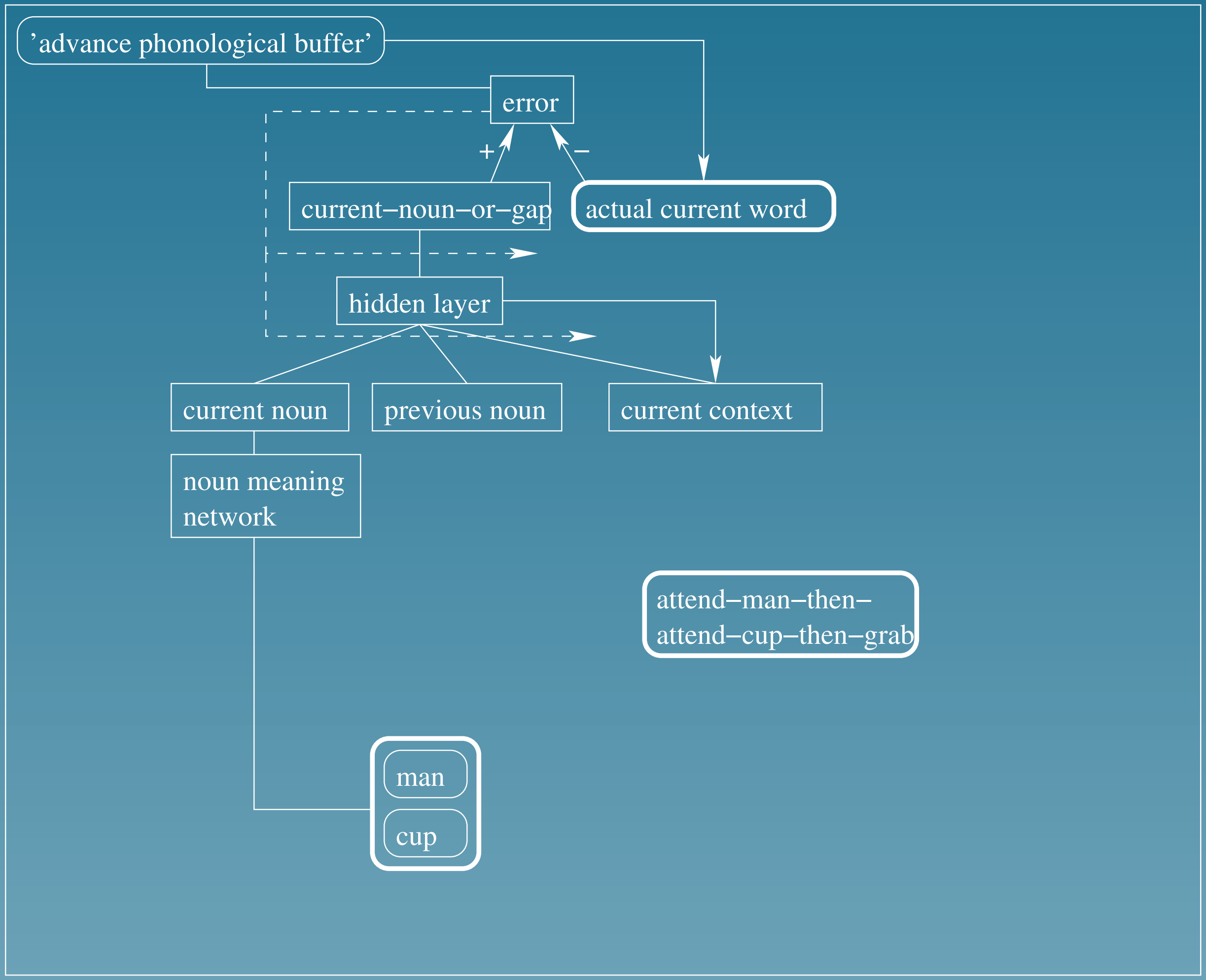
Training data (for a SVO language):

SM sequence	MAN	CUP	MAN	CUP
Word sequence	man cup			

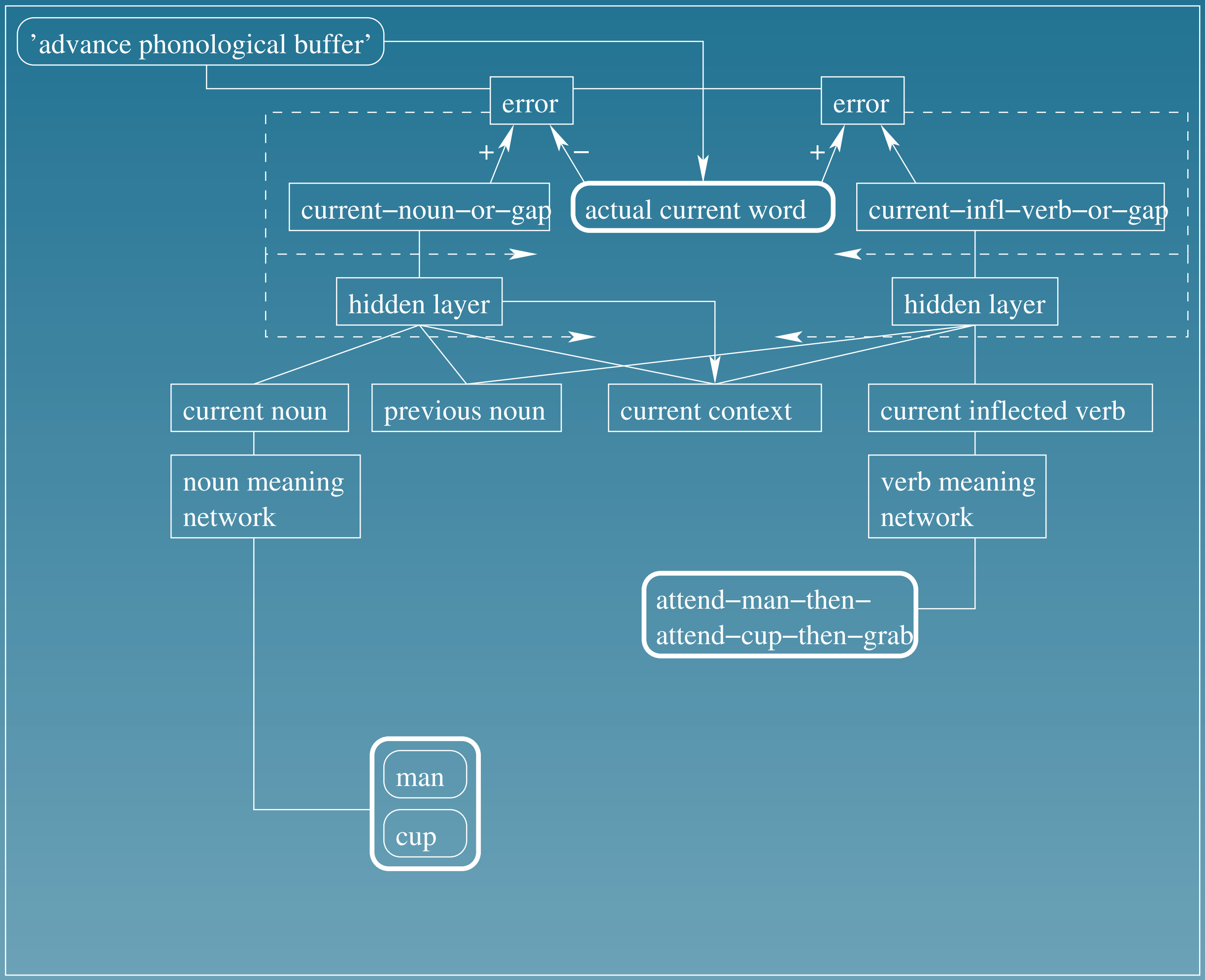
Network output after training:

SM sequence	MAN	CUP	MAN	CUP
Word sequence	man	GAP	GAP	cup

Verb sequencing network



Verb sequencing network



Results of the complete network

Training data (for a SVO language):

SM sequence	MAN/GRAB-PLAN	CUP/GRAB-PLAN	MAN/GRAB-PLAN	CUP/GRAB-PLAN
Word sequence	man grabbed cup			

Network output after training:

SM sequence	MAN/GRAB-PLAN	CUP/GRAB-PLAN	MAN/GRAB-PLAN	CUP/GRAB-PLAN
Word sequence	man/GAP	GAP/GAP	GAP/grabbed	cup/GAP

Training data from different languages results in the right kinds of gapping being learned.

Conclusions

Conclusions

Modelling sentence LFs as SM *sequences* permits some interesting new sentence-processing architectures.

Conclusions

Modelling sentence LFs as SM *sequences* permits some interesting new sentence-processing architectures.

There are some interesting similarities between a sensorimotor model of transitive actions and a Minimalist model of transitive clauses.

Conclusions

Modelling sentence LFs as SM *sequences* permits some interesting new sentence-processing architectures.

There are some interesting similarities between a sensorimotor model of transitive actions and a Minimalist model of transitive clauses.

Maybe Chomskyan syntax is not so weird after all.