# COSC451: Artificial Intelligence

## Lecture 15: How infants learn words

Alistair Knott

Dept. of Computer Science, University of Otago

# Recap

Last lecture:

- How phonological representations are encoded in the brain
  - The idea of a mirror system for phonemes
  - The idea of phonological input and output buffers
- How words are encoded in the brain
  - As phonological units
  - As semantic units
  - As associations between phonological and semantic units
- Grammatical processing in the brain
  - Broca's aphasia and Wernicke's aphasia

# Outline of today's lecture

1. Development of phonological word representations
2. Development of object and action concepts
3. Development of knowledge of word meanings

# Outline of today's lecture

1. Development of phonological word representations
2. Development of object and action concepts
3. Development of knowledge of word meanings

# Development of phonological word representations

There's good evidence that infants can form representations of words *as commonly-occurring phonological sequences* by around 8 months (Saffran *et al.*, 1996).

- In a training phase, an infant is played a stream of phonemes, where some phoneme sequences occur more often than chance. E.g. *ga bi ro to ba di ga bi ro pa la ku to ba di*

# Development of phonological word representations

There's good evidence that infants can form representations of words *as commonly-occurring phonological sequences* by around 8 months (Saffran *et al.*, 1996).

- In a training phase, an infant is played a stream of phonemes, where some phoneme sequences occur more often than chance. E.g. *ga bi ro to ba di ga bi ro pa la ku to ba di*

# Development of phonological word representations

There's good evidence that infants can form representations of words *as commonly-occurring phonological sequences* by around 8 months (Saffran *et al.*, 1996).

- In a training phase, an infant is played a stream of phonemes, where some phoneme sequences occur more often than chance. E.g. *ga bi ro to ba di ga bi ro pa la ku to ba di*

# Development of phonological word representations

There's good evidence that infants can form representations of words *as commonly-occurring phonological sequences* by around 8 months (Saffran *et al.*, 1996).

- In a training phase, an infant is played a stream of phonemes, where some phoneme sequences occur more often than chance. E.g. *ga bi ro to ba di ga bi ro pa la ku to ba di*
- In a testing phase, infants were presented with a mixture of high-probability sequences (*to ba di*) and low-probability sequences (*di ga bi*).

# Development of phonological word representations

There's good evidence that infants can form representations of words *as commonly-occurring phonological sequences* by around 8 months (Saffran *et al.*, 1996).

- In a training phase, an infant is played a stream of phonemes, where some phoneme sequences occur more often than chance. E.g. *ga bi ro to ba di ga bi ro pa la ku to ba di*
- In a testing phase, infants were presented with a mixture of high-probability sequences (*to ba di*) and low-probability sequences (*di ga bi*).
- They showed longer listening times for the low-probablity sequences, suggesting that they were perceived as novel.

# Development of phonological word representations

There's good evidence that infants can form representations of words *as commonly-occurring phonological sequences* by around 8 months (Saffran *et al.*, 1996).

- In a training phase, an infant is played a stream of phonemes, where some phoneme sequences occur more often than chance. E.g. *ga bi ro to ba di ga bi ro pa la ku to ba di*
- In a testing phase, infants were presented with a mixture of high-probability sequences (*to ba di*) and low-probability sequences (*di ga bi*).
- They showed longer listening times for the low-probablity sequences, suggesting that they were perceived as novel.

By 9 months, infants prefer to listen to nonwords which conform to the phonological rules of the exposure language (Jusczyk *et al.*, 1994)

# Development of phonological output representations

Infants produce sounds from birth.

- Children *imitate* the speech sounds they hear from very early. E.g. vowels are imitated from around 1 month old (Kuhl and Meltzoff, 1996).
- So the mirror system for phonemes starts to be learned very early.

Infants start to use syllables ('babble') around 6 months.

- To begin with, babbling is repeating a single syllable (*bababa*).
- By around 8 months, infants can produce short sequences of varied syllables (e.g. *baga*).
- Infants begin to *imitate* syllables at the time they start to babble (see e.g. Oller, 2000).

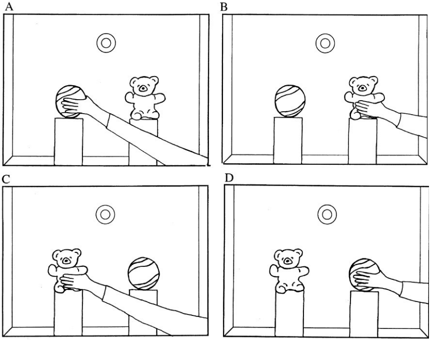
# Outline of today's lecture

1. Development of phonological word representations
2. Development of object and action concepts
3. Development of knowledge of word meanings

# Development of object and action concepts

By 8 months:

- Infants can perform reach-to-grasp actions (Dimitrijevi and Bjelakovi, 2004).
- Infants are sensitive to the intended target of a reach-to-grasp action (Woodward, 1998).



# Development of object and action concepts

By 8 months:

- Infants have a good understanding of the support/contact relations involved in grasping (Leslie, 1984).
- Infants have a good awareness of the spatiotemporal continuity of objects (Spelke *et al.*, 1994).
- Infants can recognise common objects in a variety of poses (Ruff, 1978).

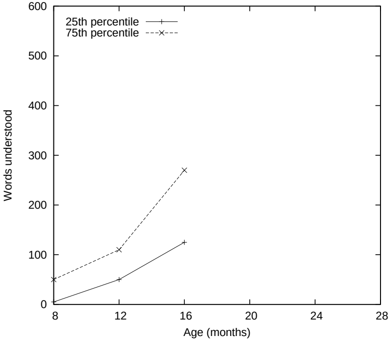
So at 8 months:

- Infants have well-developed phonological word representations;
- Infants (apparently) have well-enough developed SM concepts to serve as semantic representations of concrete words.

# Outline of today's lecture

# Development of knowledge of word meanings

Some summary data from a well-normed study of US infants (Fenson *et al.*, 1994):




- 8–12 months: infants acquire a small passive vocabulary.
- Active vocabulary really takes off at around 12 months.

# Basic associationist model of word meaning learning

The basic model:

- An infant perceives an object or action, and evokes a semantic representation.
- An adult says the word which denotes this object/action, and the infant evokes a phonological representation.
- Associations are strengthened in the infant between the semantic and phonological representations. (Straight Hebbian learning.)

This simple idea is partly right. But there are a few problems.

# Correlation problems

The stream of words infants hear is often very decorrelated from the stream of semantic representations they evoke.


(From Baldwin, 1995)

Correspondences between word forms and word meanings take the form of weak *statistical tendencies*, reflecting the fact that parents talk about things the infant is attending to 'more than chance'.

# Cross-situational learning (Siskind, 1996)

Idea:

- The infant is in a situation, where only some of his semantic concepts will be activated.
- Assume that the infant is exploring the situation, and evoking a stream of semantic concepts.
- The situation also contains mature speakers, whose utterances are more likely than chance to contain words describing situation-related concepts.
- Assume the infant is processing these utterances phonologically, and activating a stream of word forms.
- *Any word form the infant hears in this situation is more likely than chance to denote the semantic concept he is currently evoking*.

So if we treat the set of word-concept pairs as training data, we'll eventually learn word meanings.

# Refinements to cross-situational learning (1)

Cross-situational learning works. But it's *slow*. (Because situation-based correlations are only very weak.)

Problem 1: the sparseness of good training data. It's very unlikely that the child hears a word at *exactly* the time he activates the associated concept.
We can alleviate this problem if the infant can maintain a sliding window of words, and associate *each word in this window* with the currently evoked semantic concept.

A 'sliding window of words' is the phonological buffer!

- We saw last lecture that phonological short-term memory correlates with word-learning abilities. Perhaps this is why.

# Refinements to cross-situational learning (2)

There must be more clever, ‘directed’ ways of learning word meanings.

Problem (2):

- Nonhuman animals can do associative learning. But they can’t learn word meanings like human infants do.
- If human infants use simple associative learning to acquire word meanings, it’s surprising that they don’t learn *faster* from around 8 months.

Most current models of word learning assume there are mechanisms which go beyond simple associative learning.

# Social/pragmatic models of word meaning learning

An influential model by Michael Tomasello (2003) holds that infants need to acquire a set of social/pragmatic skills before they can learn words in any numbers.

There are two skills:

- Joint attention: the ability to follow an observed agent's eye gaze.
- The ability to recognise communicative intentions.

# Social/pragmatic models of word meaning learning

An influential model by Michael Tomasello (2003) holds that infants need to acquire a set of social/pragmatic skills before they can learn words in any numbers.

There are two skills:

- Joint attention: the ability to follow an observed agent's eye gaze. We have already talked about the role of joint attention in the perception of hand actions.
- The ability to recognise communicative intentions.

# Social/pragmatic models of word meaning learning

An influential model by Michael Tomasello (2003) holds that infants need to acquire a set of social/pragmatic skills before they can learn words in any numbers.

There are two skills:

- Joint attention: the ability to follow an observed agent's eye gaze. We have already talked about the role of joint attention in the perception of hand actions.
- The ability to recognise communicative intentions. We have talked about the role of intention recognition in the perception of hand actions, and articulatory actions.

# The role of joint attention in word meaning learning

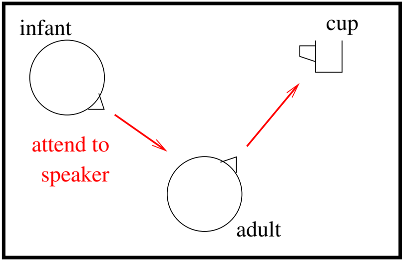
Infants can acquire language by *overhearing* mature speakers.

In some cultures, people don't talk much to children at all!
(Pinker, 1994)

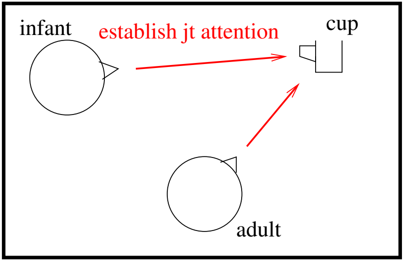
The joint attention model:

- The infant watches mature speakers talking, maybe to each other.
- While an observed speaker makes an utterance, the infant follows the speaker's gaze, and evokes semantic representations of *what the speaker is attending to*.
  (Assumption: speakers often talk about what they're attending to.)
- The infant associates *these* semantic reps with the speaker's words.

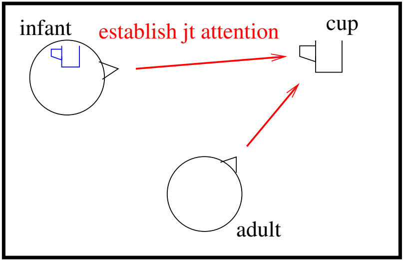
# The joint attention word-learning scenario



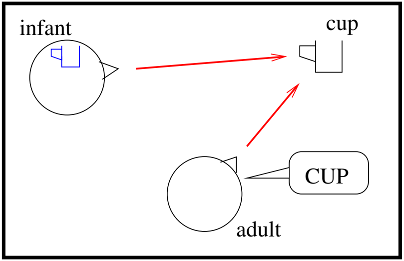
# The joint attention word-learning scenario



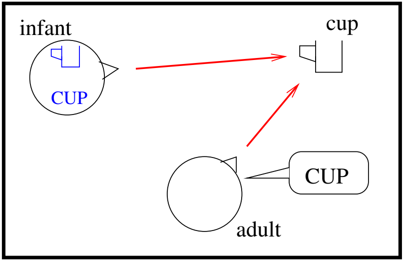
# The joint attention word-learning scenario



# The joint attention word-learning scenario



# The joint attention word-learning scenario



# Development of the joint attention capacity

Infants' ablility to establish joint attention develops gradually.

- In simple contexts (where the angle the infant must traverse to establish joint attention is small), the ability is reliable around 10 months (Scaife and Bruner, 1975).
- In more complex contexts (where the angle is larger), it doesn't become reliable until around 18 months (Butterworth and Jarrett, 1991).

# Evidence for the role of joint attention in word learning

Infants' ability to establish joint attention at 12 and 18 months is predictive of their language ability at 24 months (Mundy *et al.*, 2007).

Baldwin (1991, 1993):

- Infants will follow the gaze of an observed speaker to determine the referent of a new word the speaker is using.

# Evidence for the role of joint attention in word learning

Infants' ability to establish joint attention at 12 and 18 months is predictive of their language ability at 24 months (Mundy *et al.*, 2007).

Baldwin (1991, 1993):

- Infants will follow the gaze of an observed speaker to determine the referent of a new word the speaker is using.

Baldwin (1994): for 18-20-month-olds, simple correlation of words and evoked concepts isn't enough.

-

# Evidence for the role of joint attention in word learning

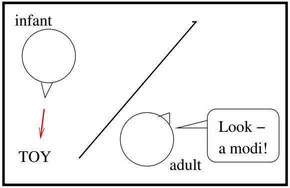
Infants' ability to establish joint attention at 12 and 18 months is predictive of their language ability at 24 months (Mundy *et al.*, 2007).

Baldwin (1991, 1993):

- Infants will follow the gaze of an observed speaker to determine the referent of a new word the speaker is using.

Baldwin (1994): for 18-20-month-olds, simple correlation of words and evoked concepts isn't enough.

- Bad learning:



# Evidence for the role of joint attention in word learning

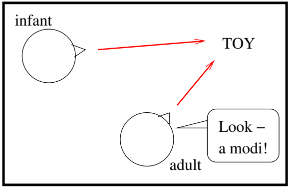
Infants' ability to establish joint attention at 12 and 18 months is predictive of their language ability at 24 months (Mundy *et al.*, 2007).

Baldwin (1991, 1993):

- Infants will follow the gaze of an observed speaker to determine the referent of a new word the speaker is using.

Baldwin (1994): for 18-20-month-olds, simple correlation of words and evoked concepts isn't enough.

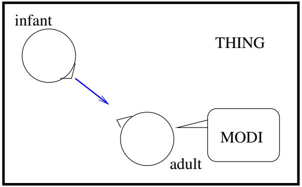
- Good learning:



# The role of communicative intention recognition in word learning

Infants don't only need to learn what words mean.
They also need to learn *what communication is*.

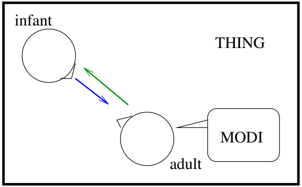
> *'Communication is where Agent1 gets Agent2 to attend to Thing, by performing a special type of action (an utterance).'*



# The role of communicative intention recognition in word learning

Infants don't only need to learn what words mean.
They also need to learn *what communication is*.

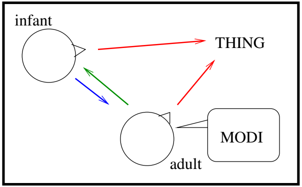
> *'Communication is where Agent1 gets Agent2 to attend to Thing, by performing a special type of action (an utterance).'*



# The role of communicative intention recognition in word learning

Infants don't only need to learn what words mean.
They also need to learn *what communication is*.

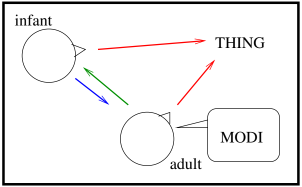
> *'Communication is where Agent1 gets Agent2 to attend to Thing, by performing a special type of action (an utterance).'*



# The role of communicative intention recognition in word learning

Infants don't only need to learn what words mean.
They also need to learn *what communication is*.

> *'Communication is where Agent1 gets Agent2 to attend to Thing, by performing a special type of action (an utterance).'*

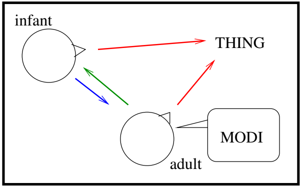



The infant has to learn that words have an attention-directing function.

# The role of communicative intention recognition in word learning

Infants don't only need to learn what words mean.
They also need to learn *what communication is*.

> *'Communication is where Agent1 gets Agent2 to attend to Thing, by performing a special type of action (an utterance).'*

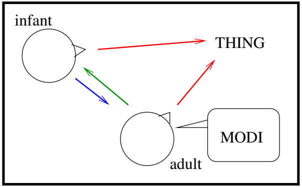



The infant has to learn that words have an attention-directing function.
Joint attention is important in learning this.

# The role of communicative intention recognition in word learning

Infants don't only need to learn what words mean.
They also need to learn *what communication is*.

> *'Communication is where Agent1 gets Agent2 to attend to Thing, by performing a special type of action (an utterance).'*

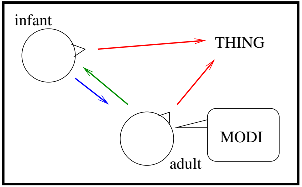



The infant has to learn that words have an attention-directing function.
Training: joint attention regularly co-occurs with observed utterances.

# The role of communicative intention recognition in word learning

Infants don't only need to learn what words mean.
They also need to learn *what communication is*.

> *'Communication is where Agent1 gets Agent2 to attend to Thing, by performing a special type of action (an utterance).'*




The infant has to learn that words have an attention-directing function.
After training, utterances actively direct an infant's attention.

# Developing the concept of communicative intentions

Infants begin following pointing gestures around 11–12 months (Butterworth, 2006; Carpenter *et al.*, 1998).

Behne *et al.* (2005): infants had the task of looking for a hidden object.

- The experimenter helped them by giving them nonverbal ‘communicative gestures’ (alternately looking at the infant and the hiding place, with and without pointing at the hiding place).
- These gestures began to be understood at 14 months.
- ‘Non-communicative’ looking / pointing was not helpful.

# Nonhuman primates and communicative intention recognition

We know that monkeys/great apes *can* recognise the intentions of other agents. And we know they *can* follow gaze.

Interestingly, great apes don't seem able to recognise *specifically communicative* intentions (Call *et al.*, 2004).

- E.g. They can't make use of 'communicative gestures' (looking or pointing) in a hide-and-seek game.

Tomasello: nonhuman primates can't represent communicative gestures. This is why they don't have language.

# Some questions about Tomasello's model of word learning

Two questions:

1. How are communicative intentions represented in the brain? (Tomasello doesn't say.)
   They involve 'things in the world' (the speaker, the hearer) and 'things the utterance describes' (the cup).
   How to represent these very different types of thing?
   E.g. 'Agent1 wants me to notice the cup'
   E.g. 'Agent1 wants me to notice the man grabbed the cup'
2. What's the *mechanism* via which communicative intention recognition supports word meaning learning?

# Representing events and intentions

We've already got some ideas about how events and intentions are represented in the brain.

- Events involving actions (e.g. *The man grabbed a cup*) are represented (in working memory) as SM sequence plans in PFC.
- (So events are represented *as intentions*.)

So according to our model, communicative actions are probably represented in PFC somehow.

The question: what's the difference between perceiving a regular physical action, and a communicative action?

# An idea about how to represent communicative actions

If we buy the idea that semantic representations are *sequential*, then here's one way we could represent communicative actions.

We first recognise a communicative action as a regular motor action.

- We attend to the agent (speaker), then to the patient (hearer), and then classify the action (communicate).

Recognising a *communicate* action has two special side-effects.

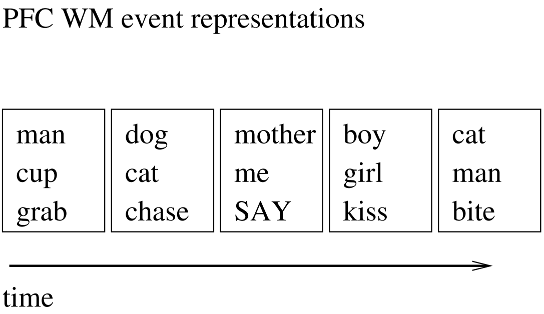
- We move into a special 'verbal mode', where links between words and concepts are turned on.
- We clear our working memory 'event' buffer.

Any words we now hear will create a brand new semantic rep.

In this model, the special relation between an utterance and its content is one of temporal adjacency.

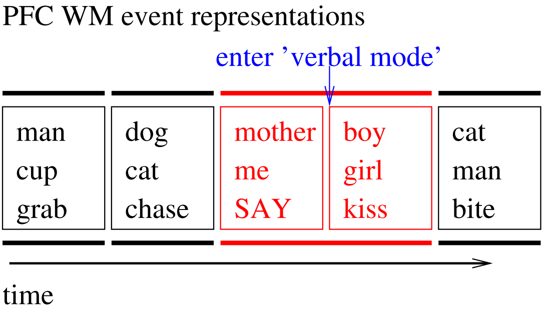
# PFC representation of communicative actions

We can imagine a stream of successive WM event representations in PFC.



# PFC representation of communicative actions

We can imagine a stream of successive WM event representations in PFC.




Communicative events are special, because they comprise a *pair* of successive event representations, with a switch to 'verbal mode' in between.

# An idea about how communicative intention recognition supports word learning

Recall: the stream of words infants hear is very decorrelated with the stream of semantic representations they evoke.

Infant's semantic representations

| | | | | |
|---|---|---|---|---|
| man | dog | mother | boy | cat |
| cup | cat | me | girl | man |
| grab | chase | SAY | kiss | bite |

Words the infant hears

| | | | | |
|---|---|---|---|---|
| office | John | | boy | mastodon |
| boss | beer | | girl | museum |
| cretin | like | | kiss | at |

time

# An idea about how communicative intention recognition supports word learning

Recall: the stream of words infants hear is very decorrelated with the stream of semantic representations they evoke.

Infant's semantic representations

| | | | | |
|---|---|---|---|---|
| man<br>cup<br>grab | dog<br>cat<br>chase | mother<br>me<br>SAY | boy<br>girl<br>kiss | cat<br>man<br>bite |

Words the infant hears

| | | | | |
|---|---|---|---|---|
| office<br>boss<br>cretin | John<br>beer<br>like | | boy<br>girl<br>kiss | mastodon<br>museum<br>at |

time →

But once the child has learned to establish *joint attention* with observed adults, there are some special moments where the two streams become well correlated.

# An idea about how communicative intention recognition supports word learning

Recall: the stream of words infants hear is very decorrelated with the stream of semantic representations they evoke.

Infant's semantic representations

| | | | | |
|---|---|---|---|---|
| man | dog | mother | boy | cat |
| cup | cat | me | girl | man |
| grab | chase | SAY | kiss | bite |

Words the infant hears

| | | | |
|---|---|---|---|
| office | John | boy | mastodon |
| boss | beer | girl | museum |
| cretin | like | kiss | at |

time

These moments occur right after a *SAY* action has been observed.

# An idea about how communicative intention recognition supports word learning

Idea:

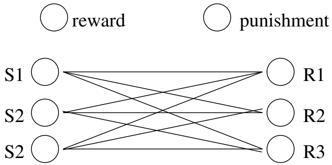
- ‘Entering verbal mode’ is an action, which the infant can choose to do at any time.
- In verbal mode, if you can predict the next word you hear, you get a reward—otherwise you get a punishment.

In this scenario, the infant can *learn by reinforcement* to enter verbal mode after perceiving a communicative action.

- (The infant has to have a small vocabulary to kickstart the process.)

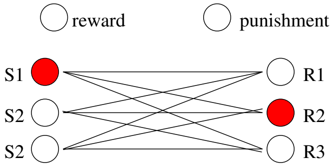
# Reinforcement learning for entering verbal mode

Here's the basic reinforcement learning paradigm.



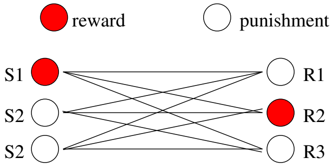
# Reinforcement learning for entering verbal mode

Here's the basic reinforcement learning paradigm.



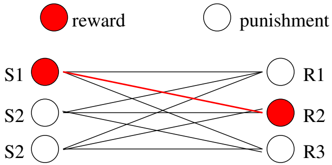
# Reinforcement learning for entering verbal mode

Here's the basic reinforcement learning paradigm.



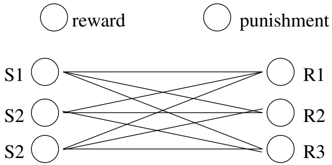
# Reinforcement learning for entering verbal mode

Here's the basic reinforcement learning paradigm.



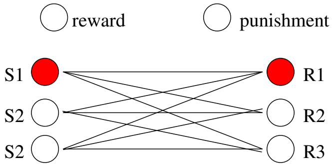
# Reinforcement learning for entering verbal mode

Here's the basic reinforcement learning paradigm.



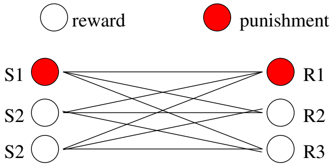
# Reinforcement learning for entering verbal mode

Here's the basic reinforcement learning paradigm.



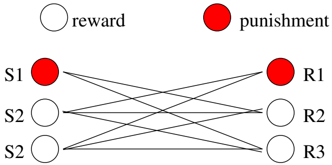
# Reinforcement learning for entering verbal mode

Here's the basic reinforcement learning paradigm.



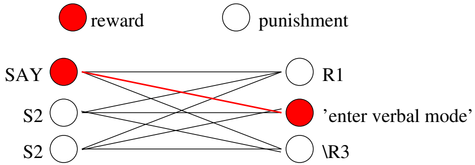
# Reinforcement learning for entering verbal mode

Here's the basic reinforcement learning paradigm.



# Reinforcement learning for entering verbal mode

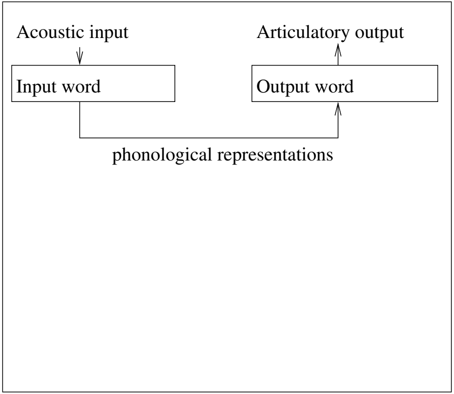
Here's how it would look for learning to enter verbal mode.




(Provided you get a reward for correctly-predicted word.)

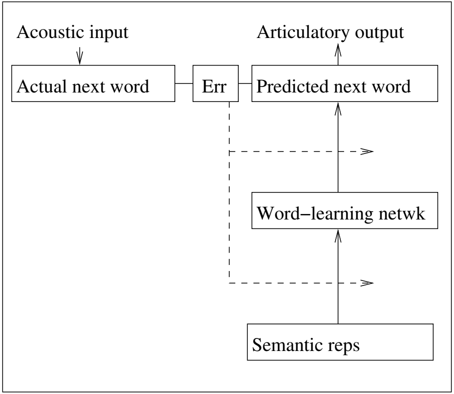
# The network for learning word semantics

Here's an extract from the phonological circuit I showed last time.



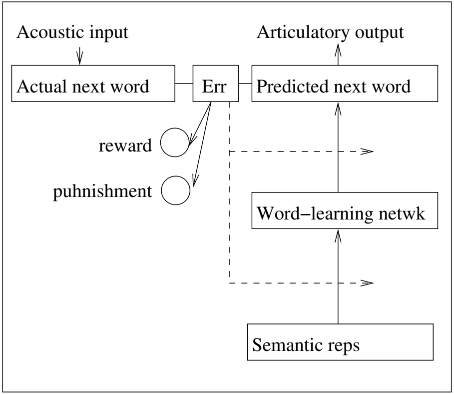
# The network for learning word semantics

To learn words, we must learn how to *reproduce* an input word from a semantic rep.



# The network for learning word semantics

To learn words, we must learn how to *reproduce* an input word from a semantic rep.



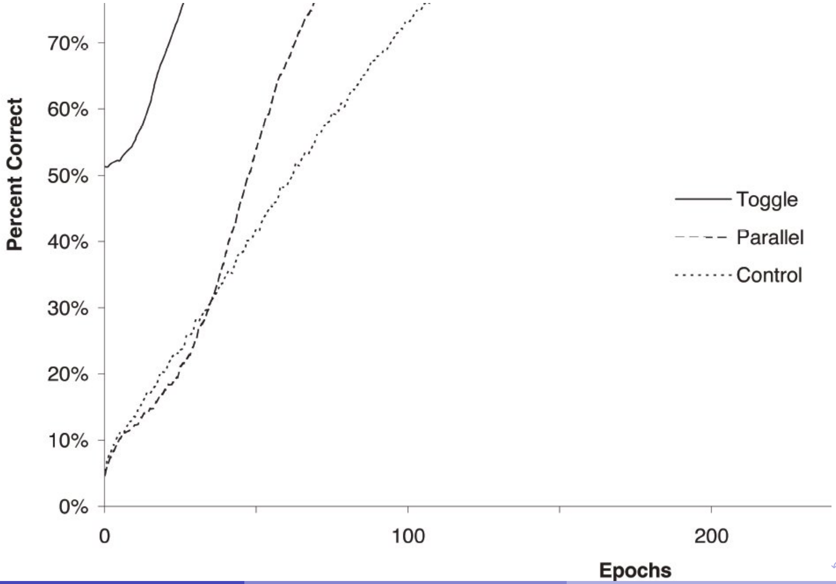

The error term can generate an ‘internal’ reward signal.

# Pragmatic bootstrapping

Here's a model where word-learning and communicative action recognition bootstrap one other.

- The infant begins by learning a small vocabulary (inefficiently) using cross-situational learning.
  (No social/pragmatic skills needed here.)
- During development, the infant randomly experiments with entering verbal mode.
- Initially, it's never a good idea.
- After a small vocabulary is learned, it's sometimes a good idea, so the child tries it more often.
- This makes word learning more efficient.

# A simulation (done by Greg Caza)



# Summary

How infants learn words (empirical data):

- Preliminaries: word-sized phonological representations and concepts
- A small passive vocabulary
- Fast expansion of passive and active vocabulary size, perhaps correlated with development of social/pragmatic skills

Computational models of early word learning:

- Cross-situational learning is probably involved
- Joint attention is probably involved
- Perhaps vocabulary development bootstraps development of the concept of a communicative actions