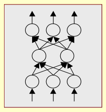
## Modelling Language Development



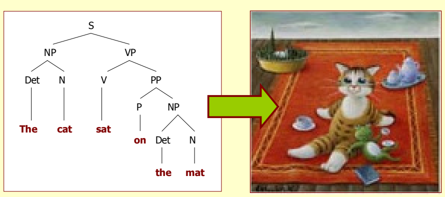
*Fiona M Richardson*
Developmental Neurocognition Lab
Birkbeck, University of London

# Overview

- Language development
- Connectionism
- Models of language development in connectionism
- Summary
- Take home message

## Language Development

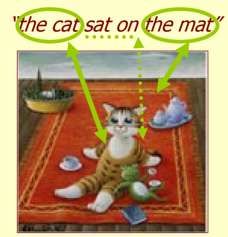
- Children acquire a specific language

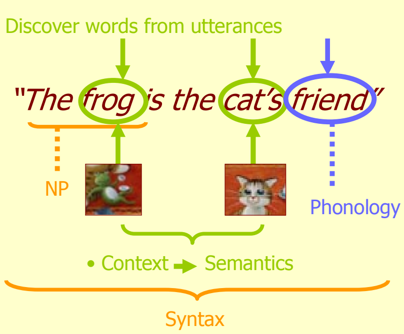



Mapping **form** to **meaning**

## Discovering Mappings

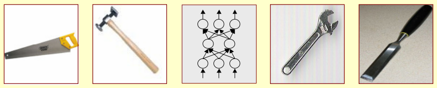
- Segment speech into meaning relevant chunks
- Segment the scene into speech-chunk relevant meanings




## How might connectionist models help us to understand language development?

# Models as Tools



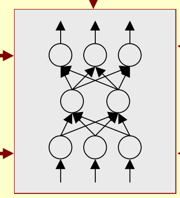
- Parameterise, develop, and test theories
- Try to explain *why*
- Controlled means of testing
- Compare with empirical data
- Generate predictions

# Properties of connectionist models

**Ability to learn**

Complex tasks can be performed by simple units

Content addressable memory

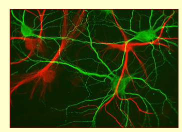


Generalise from experience

Resistant to damage

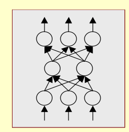
# Connectionist models are models of learning, which we can use to explore the mechanisms of developmental change

# Where's the knowledge?



Hippocampal neurons

*(with glial cells shown in red)*



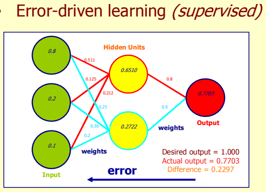
Representation of a connectionist network

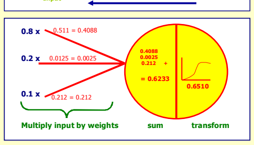
- neurons = *units*
- connections = *weights*

**Knowledge is stored in the weights and is acquired through learning**

# Learning in connectionism: back-propagation

- Error-driven learning *(supervised)*




(1) Calculate unit activations

(2) Find the difference between desired and actual output

(3) Back-propagate this error down the network

*Adjust connection weights to reduce the error at output*

* **Learning rate**: proportion of the weight change (range 0 to 1)

* **Momentum**: scales the extent to which a previous weight change carries through to the current weight change

# Models

- English past tense formation
  - Plunkett & Marchman (1993)
- Word recognition and naming
  - Seidenberg & McClelland (1989)
- The mental lexicon
  - Elman (1990, 2004)
- Lexical development
  - Li, Farkas & MacWhinney (2004)

## English Past Tense Formation
## Plunkett & Marchman (1993)

## The English Past Tense

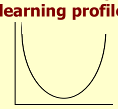
- Regulars: + 'ed'
  i.e. talked, walked, baked, parked etc …
- Irregulars:

| | | |
|---|---|---|
| *Arbitrary* | go | → went |
| *No change* | hit | → hit |
| *Vowel change* | come | → came |
| *Blend* | creep | → crept |

## Learning the English Past Tense

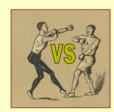
- Children initially use the correct past tense for a limited set of high-frequency irregular items
  - i.e. go → went
- Then they start to make errors: } **Over-regularisation**
  - i.e [go → goed], [hit → hitted]

**"micro" U-shaped learning profile**

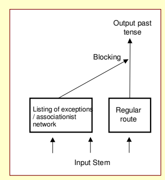


As children experience more verb forms, they discover the regular pattern of the past-tense

## The Debate

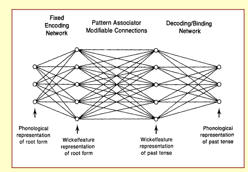



**Dual Route**




(Pinker, 1991, 1994)

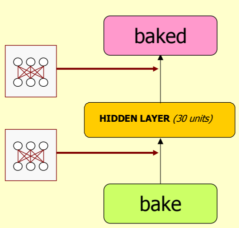
**Associative account**




(Rumelhart & McClelland, 1986)

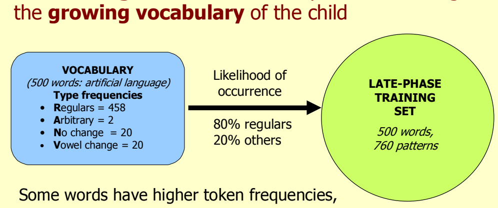
## The Plunkett & Marchman model

- A simple feed-forward connectionist network




## Training

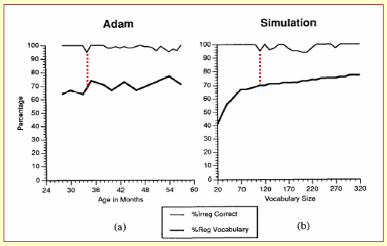
- Model trained using back-propagation
- The **training set** of the model expands, emulating the **growing vocabulary** of the child




Some words have higher token frequencies, so they occur more times in the training set

## Model Data

• Onset of over-regularisation errors at similar point




- Increase in the proportion of regular verbs triggers a shift from rote learning to generalisation

## What the model shows us…

- Exhibits over-regularisation
- Irregular verbs with a high token frequency are less prone to over-regularisation errors

*"A single mechanism learning system may offer an alternative account of the transition from rote learning process to system building"*

(Plunkett & Marchman, 1993 p58)

## Word Recognition and Naming
## Seidenberg & McClelland (1989)

## Learning to Read

- Reading: learn how *spoken* forms **map** onto unfamiliar *written* forms
- Phonics: "sounding out" words
  - i.e.


**Match the pronunciation of a written word to a known phonological form**

- The pronunciation of words is generally systematic but there are **inconsistencies:**
  - Grapheme-phoneme correspondences [i.e. CAVE, GAVE, SAVE …. HAVE]
  - Syllables on orthography [i.e. BAKED – NAKED]
  - Morphology [i.e. PREVIEW, DECODE / DELIVER, PRETENSE]

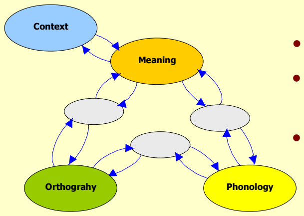
**A quasi-regular system**

## Development of Reading Skills

**Strengthen knowledge of spelling-sound correspondences**

- Younger less-skilled readers:
  - Take longer to name words than older readers
  - Have more difficulty with words associated with multiple pronunciations
  - Show larger regularity effects than older readers
- Older more-skilled readers:
  - Differences only persist for low frequency items

## Reading Words

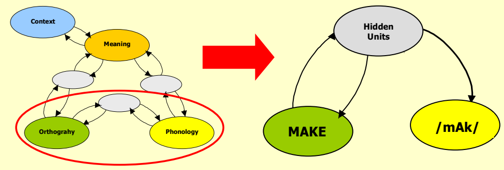



- 3 types of codes
- Distributed representations
- Interactive processing

(Seidenberg & McClelland, 1989)

A minimal model of lexical processing, where as much as possible is left to the mechanisms of **learning**

## Seidenberg & McClelland's model



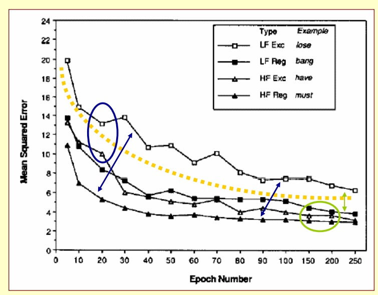

- No lexicon
- No pronunciation rules

} **A single processing mechanism**

## Training the Model

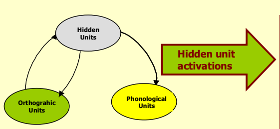
- The model was trained using back-propagation
- Training set:
  - 2897 monosyllabic words consisting of three or more letters from the Kucera & Francis (1967) word count
  - The probability of the word being presented to the model was related frequency of occurrence

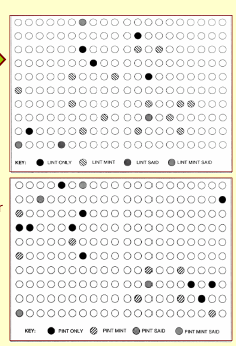
## Model Data




- **Gradual improvement in reading skill**
- **Early in learning:**
  - pronounced difficulty with exception pronunciations
  - higher regularity effects
- **Later in learning:**
  - similar performance across word classes
  - differences persist only for low frequency exceptions

## Taking a look inside...



- Similarly spelled rhymes activate the largest number of common units (LINT/MINT = 14)

- Similarly spelled non-rhymes activate a smaller number of common units (PINT/MINT = 18)

- Unrelated words activate a very small number of common units (LINT/SAID & PINT/SAID = 1)

**Reflect generalisations concerning regularities in the lexicon**

## What the model shows us...

- Shows the developmental course of acquisition
- Model captures key aspects of child data and differences in reading skill
- Claim regarding representation of orthographic knowledge:
  - *More congruent with knowledge distributed across connection weights than with pronunciation rules*
- Model can be used to explore reading difficulties
  - i.e. dylsexia

## The Mental Lexicon
## Elman (1990, 2004)

## **Elman:** the mental lexicon

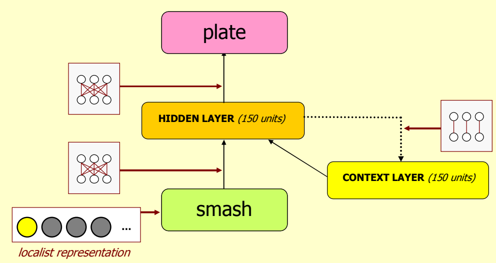
- Words as cues to meaning
- Act directly on mental states

*"It is in the precise nature of their causal effects the specific properties of words - phonological, syntactic, semantic and so forth - are revealed."*

(Elman, 2004 p301)

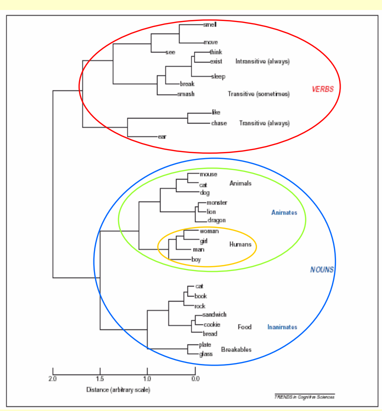
## Elman's model

- A Simple Recurrent Network (**SRN**)
- **Task:** predict next word in the sequence
  - Used a corpus of 10,000 two and three-word sentences generated by an artificial but natural-like grammar
  - Learn about the abstract relationships between constituents, rather than memorising the corpus
  - Generate expectancies about grammatically possible successors
  - Explore the internal structure of the model

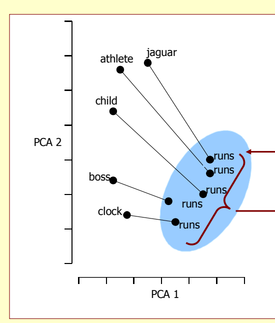



**Sequence prediction task**: Woman smash plate

## Internal structure




## Context sensitivity




- The meaning of an individual word can differ depending on context
- Get integration *for free*

## What the model shows us…

- Emergence of structure
  - Categories
  - Similarity
- Context-sensitivity
  - Integration *for free*
- No lexicon in the usual sense
  - lexical knowledge is implicit in the effects that words have on internal states

## Lexical Development
**Li, Farkas & MacWhinney (2004)**

## The Developing Lexicon

- Vocabulary grows

**Vocabulary spurt 18-20 months**

**confusion**

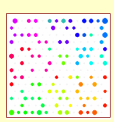
"bowl"

- Children do not receive constant feedback on what is incorrect
- Effects of age of acquisition (AoA)
  - Faster at reading and naming words acquired early in comparison to those acquired later

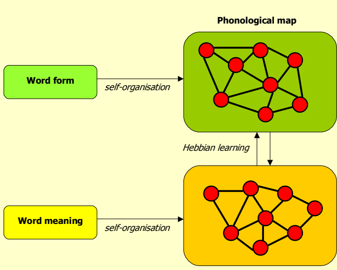
## Problems with previous models…

- Neither the **vocabulary** nor the **model** grow
  … but vocabulary grows
- Artificial lexicon
  … can use corpus-based speech data
- Use of supervised learning
  … is teacher-driven learning really appropriate?
  … what about *catastrophic interference?*
  (over-writing of previous learning)

## The DevLex Model



topographic similarity clustering

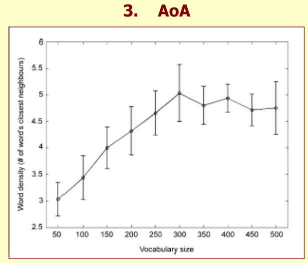



**GMAPS**

*Growing self-organising maps*

## Results

**3. AoA**



## What the model shows us…

- Using a growing lexicon with real language and unsupervised learning:
  - Emergent model of lexical development
  - A model with similarity-based clustering can account for:
    - Development of lexical representations
    - AoA effects (competition)
    - Confusion (organisation)

## Summary

- Language: mapping **form** to **meaning**
- Connectionist models are **learning** models
- Development and **emergence**
  - It's not just the end-state
- Models can **capture** developmental **data**
  - Principles at work, why certain phenomena occur
- Offer alternatives

## Take home message

Connectionist models are well-suited to exploring the process of language development – allowing us to explore the emergence of systems used in the comprehension and production of language. They have already offered both insights and alternatives in theory-building, and have the potential to continue doing so.

## End of Talk

*Thank you for listening*

**Selected References:**


- Chiat, S. (2001). Mapping theories of developmental language impairment: premises predictions and evidence. In, D. Bishop (Ed.). *Language and cognitive processes in developmental disorders.* Psychology Press Ltd: Hove, East Sussex, UK.
- Elman, J.L. (1990). Finding structure in time. *Cognitive Science*, 14, 179-211
- Elman, J.L. (2004). An alternative view of the mental lexicon. *TICS*, 8 (7), 301-306
- Li, P., Farkas, I., & MacWhinney, B. (2004). Early lexical development in a self-organizing neural network. *Neural Networks*, 17, 1345-1362
- McClelland, J.L., & Patterson, K. (2002). Rules or connections in past-tense inflections: what does the evidence rule out? *TICS*, 6 (11), 465-472
- Pinker, S., & Ullman, M.T. (2002). The past and future of the past tense. *TICS*, 6 (11) 456-463
- Plunkett, K., & Marchman, V. (1991). U-shaped learning and frequency effects in a multi-layered perceptron: Implications for child language acquisition. *Cognition,* 38, 43-102.
- Plunkett, K., & Marchman, V. (1993). From rote learning to system building: acquiring verb morphology in children and connectionist nets. *Cognition,* 48, 21-69.
- Seidenberg, M.S., & McClelland, J.L. (1989). A distributed, developmental model of word recognition and naming. *Psychological Review*, 96 (4), 523-568.