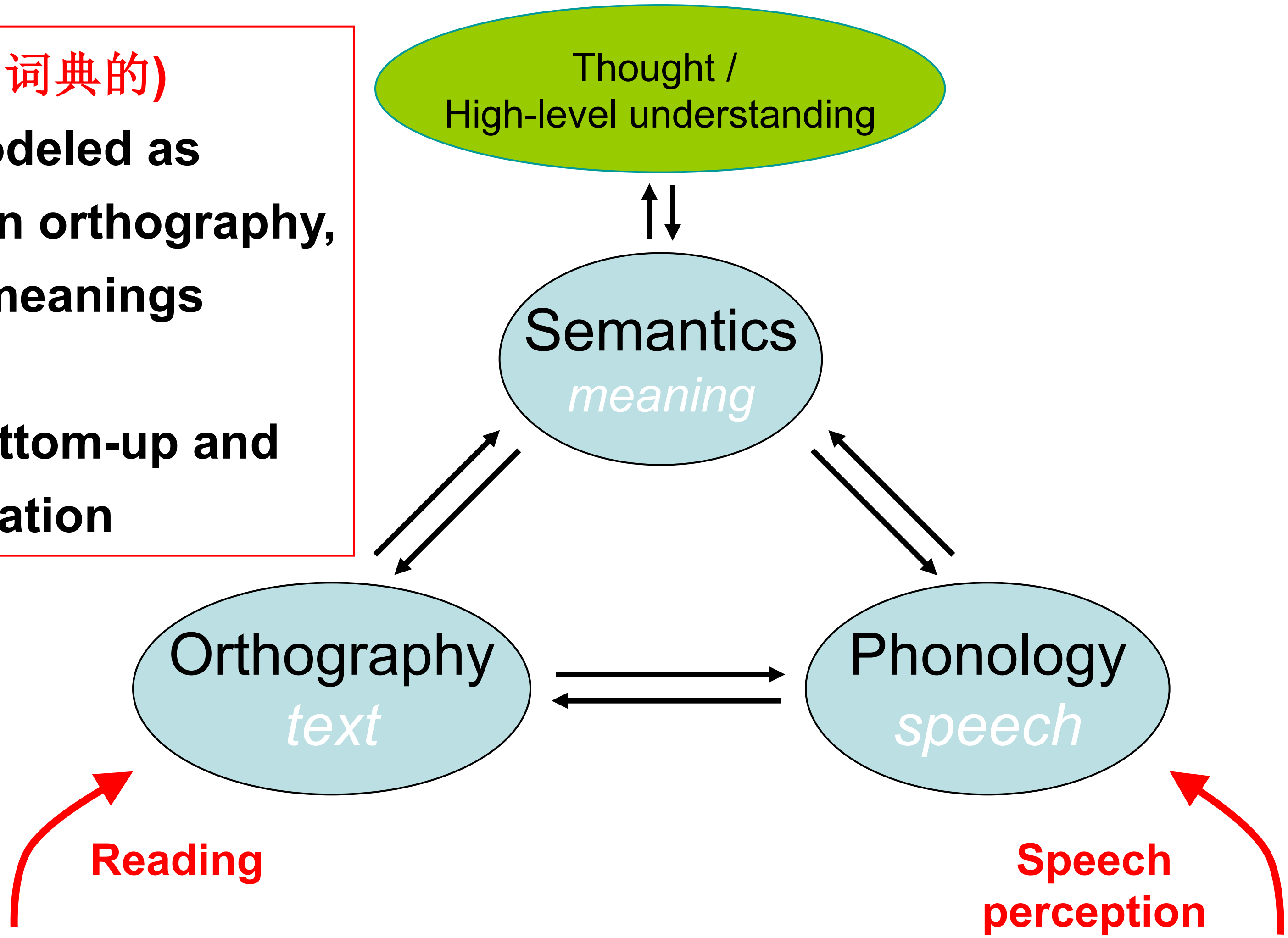
Language Comprehension

Speech Perception Semantic Processing & Naming Deficits

Triangle Model

Lexical **(词汇的，词典的)** **knowledge is modeled as mapping between orthography, phonology and meanings**

Integration of bottom-up and top-down information

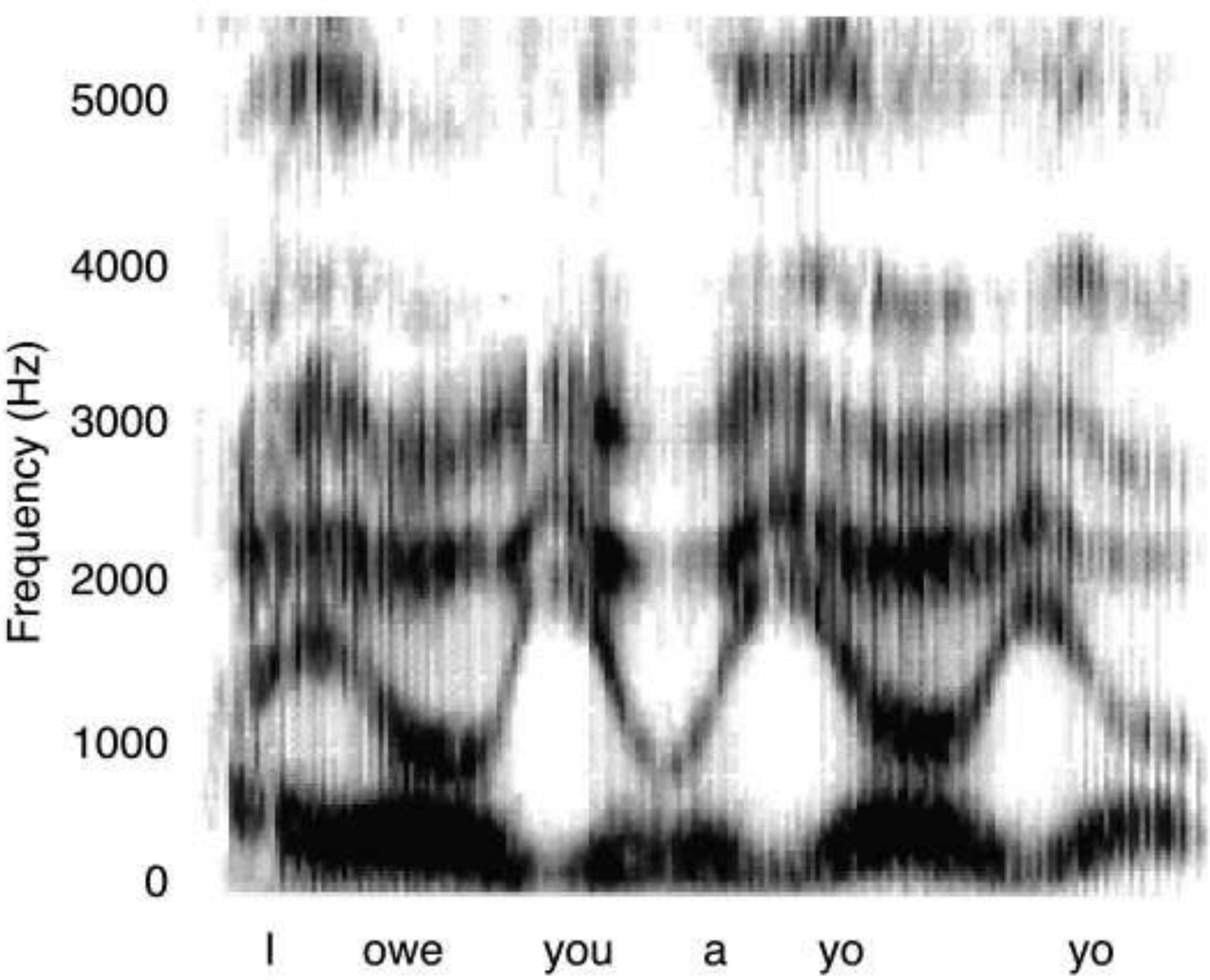


Connectionist framework for lexical processing, adapted from Seidenberg and McClelland (1989) and Plaut et al (1996).

Speech Perception

- The first step in comprehending spoken language is to identify the words being spoken, performed in multiple stages:
  - Phonemes (音素) are detected (/b/, /e/, /t/, /e/, /r/,)
  - Phonemes are combined into syllables (/be/ /ter/)
  - Syllables are combined into words (“better”)
  - Word meaning retrieved from memory

Spectrogram (声谱): I owe you a yo-yo



How many words were spoken (in Finnish)?

1) “Hyvää huomenta” (two words) → “Good morning”

2) “Kiitoksia oikein paljon” (three words) → “Thank you”

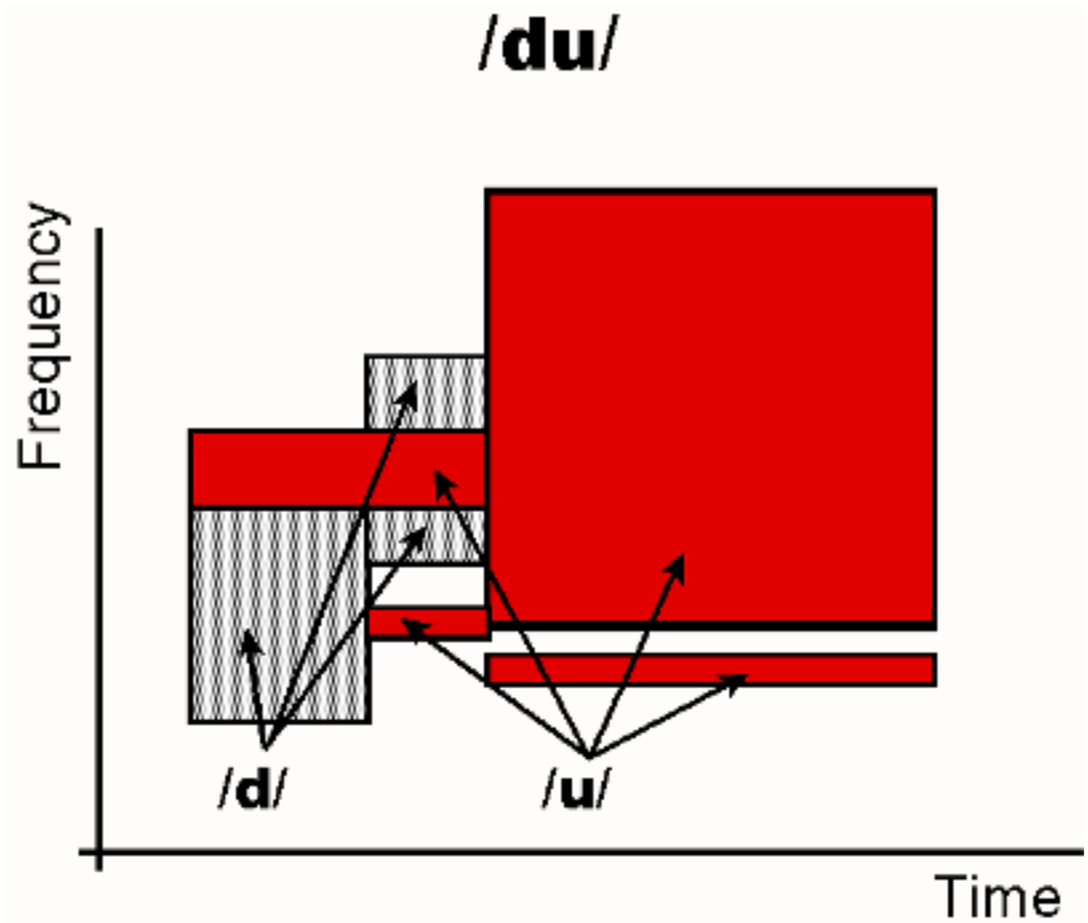
3) “Ilmatyynyalukseni on täynnä ankeriaita” (four words) → “My hovercraft is full of Eels”

Speech perception: two problems

- Words are not neatly segmented (e.g., by pauses)

- Difficult to identify phonemes
  - Coarticulation (连音) = consecutive speech sounds blend into each other due to mechanical constraints on articulators (发音器官)
  - Speaker differences; pitch (音调,音高) affected by age and sex; different dialects, talking speeds etc.

Combining the two phonemes /d/ and /u/ leads to coarticulation where the frequency characteristics associated with /d/ are blended with /u/



How Do Listeners Resolve Ambiguity in Acoustic Input?

- Use of context
  - Cross-modal context
    - ✓ E.g. use of visual cues: McGurk effect
  - Semantic context
    - ✓ E.g. phonemic restoration effect

Context is a form of top-down information

McGurk Effect

Perception of auditory event affected by visual processing

Harry McGurk and John MacDonald in "Hearing lips and seeing voices", *Nature* **264**, 746-748 (1976).

McGurk Effect

- McGurk effect in video:
  - lip movements = “ga”
  - speech sound = “ba”
  - speech perception = “da” (for 98% of adults)
- Demonstrates parallel & interactive processing: speech perception is based on multiple sources of information, e.g. lip movements, auditory information
- Brain makes reasonable assumption that both sources are informative and “fuses” the information

Effect of Semantic Context

- Pollack & Pickett (1964)
  - Recorded several conversations
  - Subjects in their experiment had to identify the words in the conversation
  - When words were spliced out (剪接出) of the conversation and then presented auditorily, subjects identified the correct word only 47% of the time
  - When context was provided, words were identified with higher accuracy
  - Clarity of speech is an illusion; we hear what we want to hear

Phonemic restoration

Auditory presentation	Perception
Legislature	legislature
Legi*lature	legislature
Legi_lature	legi lature

If a speech sound is replaced by a noise (a cough or a buzz), then listeners think they have heard the speech sound anyway. Furthermore, they cannot tell exactly where the noise was in the utterance

It was found that the *eel was on the axle.	wheel
It was found that the *eel was on the shoe.	heel
It was found that the *eel was on the orange.	peel
It was found that the *eel was on the table.	meal

Warren, R. M. (1970). Perceptual restorations of missing speech sounds. *Science*, **167**, 392-393.

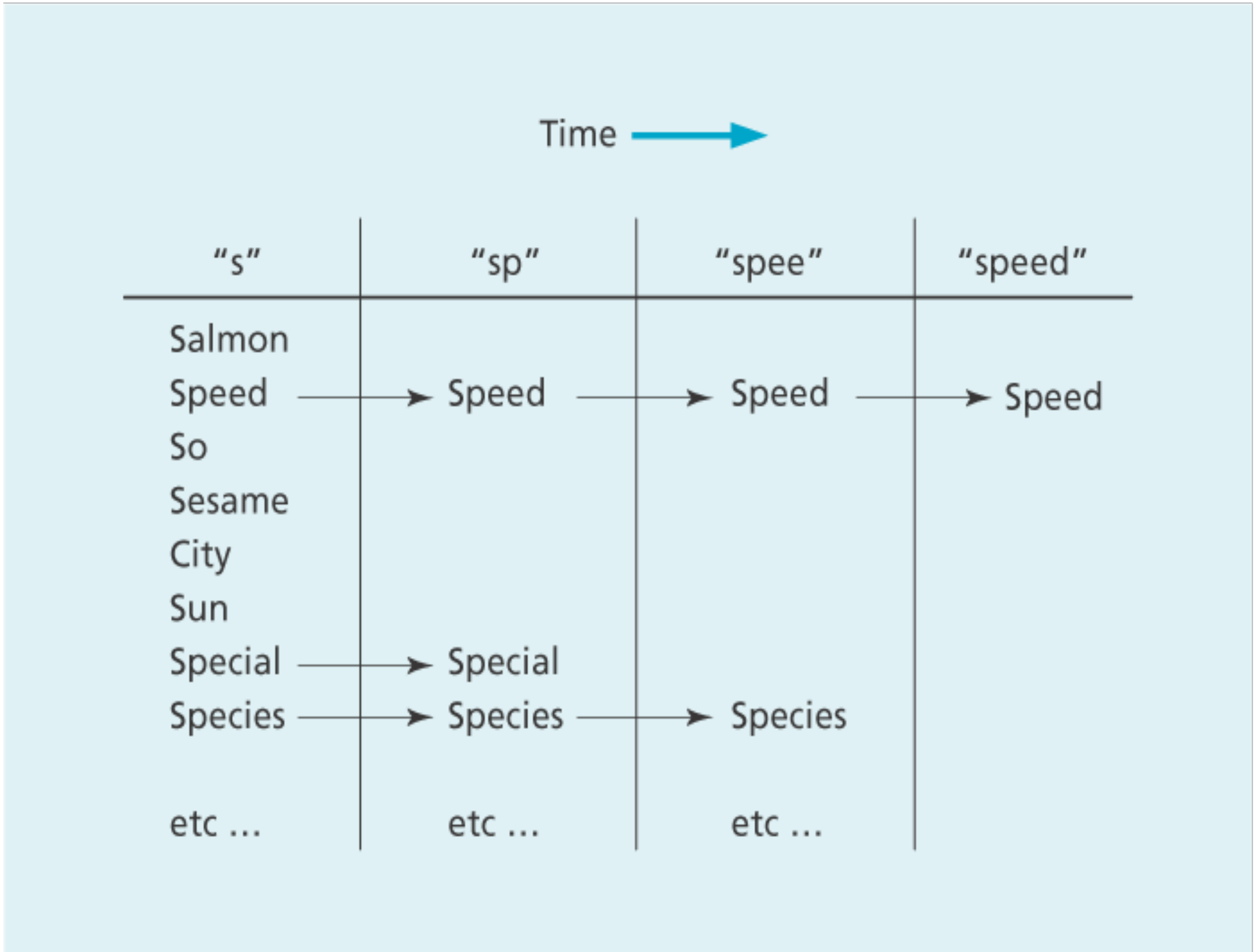
Models of Spoken Word Identification

- The Cohort Model (竞争者模型，队列模型)
  - Marslen-Wilson & Welsh, 1978
  - Revised, Marslen-Wilson, 1989

- The TRACE Model
  - Similar to the Interactive Activation model
  - McClelland & Elman, 1986

Two leading models, TRACE and Cohort Model, have much in common
Both depend on competition between partially activated candidates for the word's identity

Online word recognition: The Cohort Model



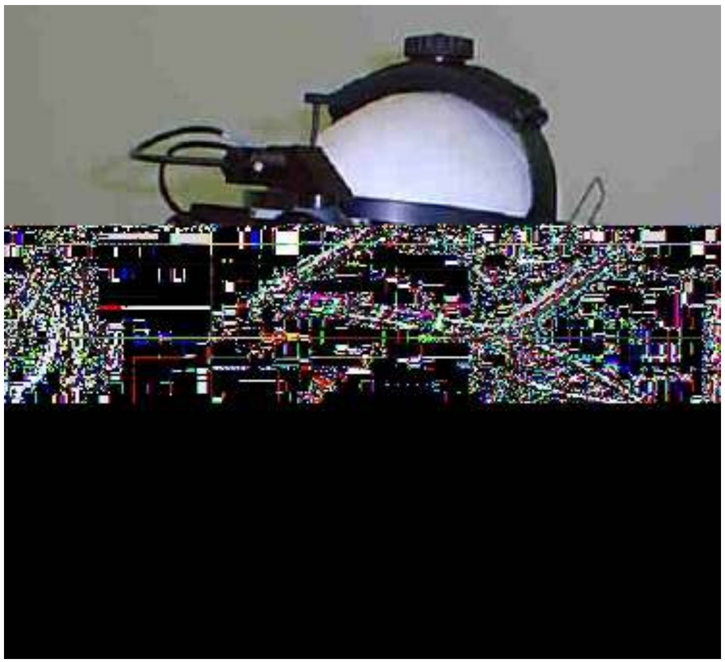
Recognizing Spoken Words: The Cohort Model

- All candidates considered in parallel
- Candidates eliminated as more evidence becomes available in the speech input
- Uniqueness point occurs when only one candidate remains

We might not be conscious of the candidate words being activated

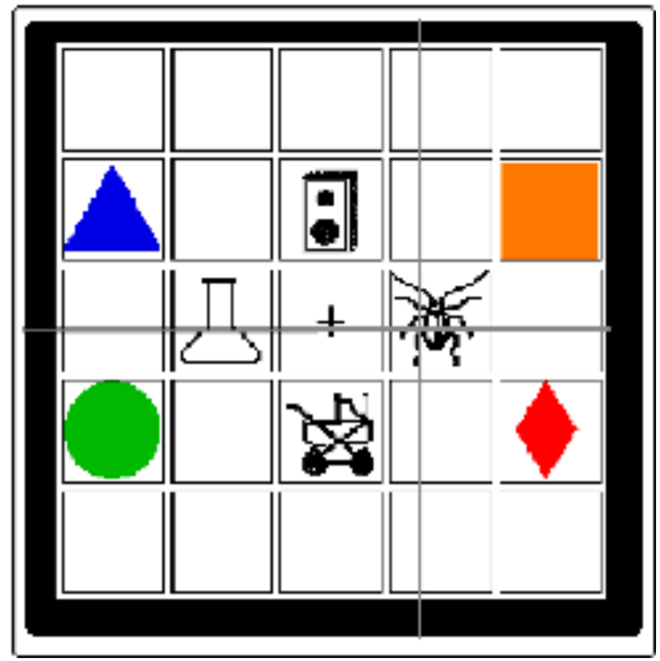
Analyzing speech perception with eye tracking

Point to the beaker



Eye tracking device to measure

where subjects are looking



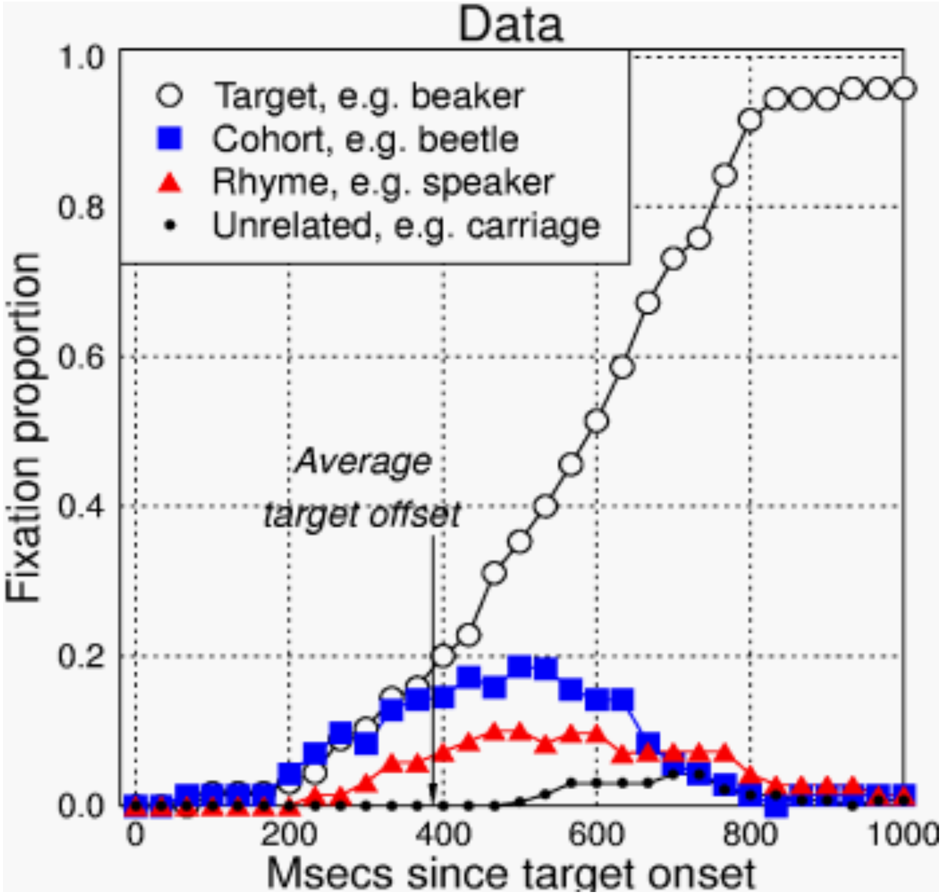
Allopenna, Magnuson & Tanenhaus (1998)

Human Eye Tracking Data

- Instructions: “Point to the …”
- Display contains:
  - Target beaker
  - And at least one of:
    - ✓ Cohort competitor beetle
    - ✓ Rhyme competitor speaker
    - ✓ Unrelated word carriage
- Participants wore head-mounted eyetracker
  - People tend to look at objects that are mentioned
  - Especially before reaching for them
  - How quickly do they look at the objects related to the target word?

Allopenna, Magnuson & Tanenhaus (1998)

Human Eye Tracking Data

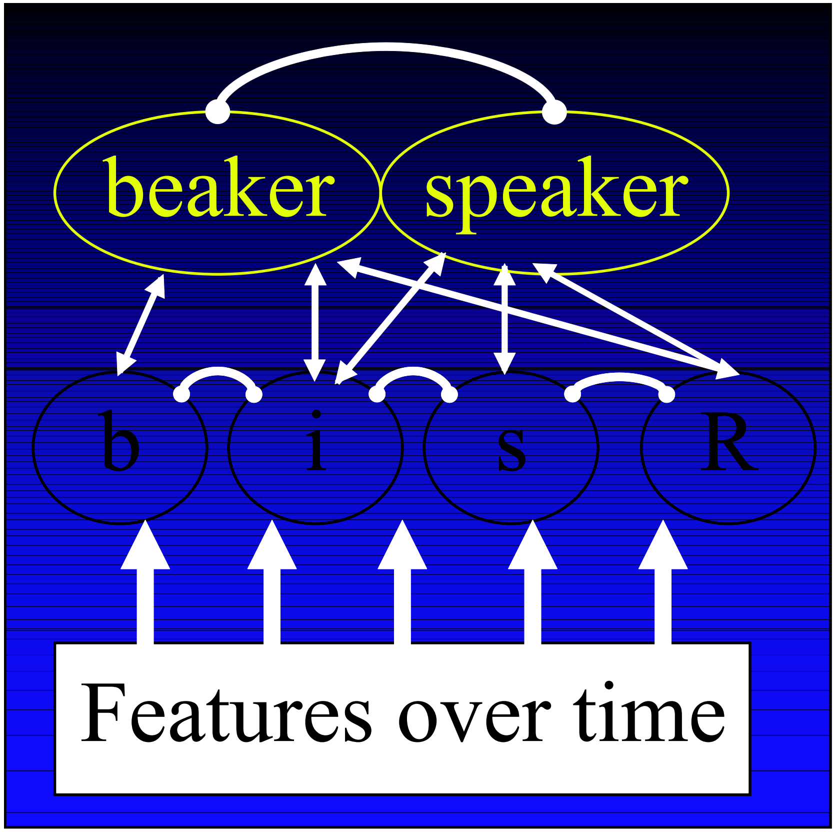


Plot shows the probability of fixating on an object as a function of time

Allopenna, Magnuson & Tanenhaus (1998)

TRACE: a neural network model

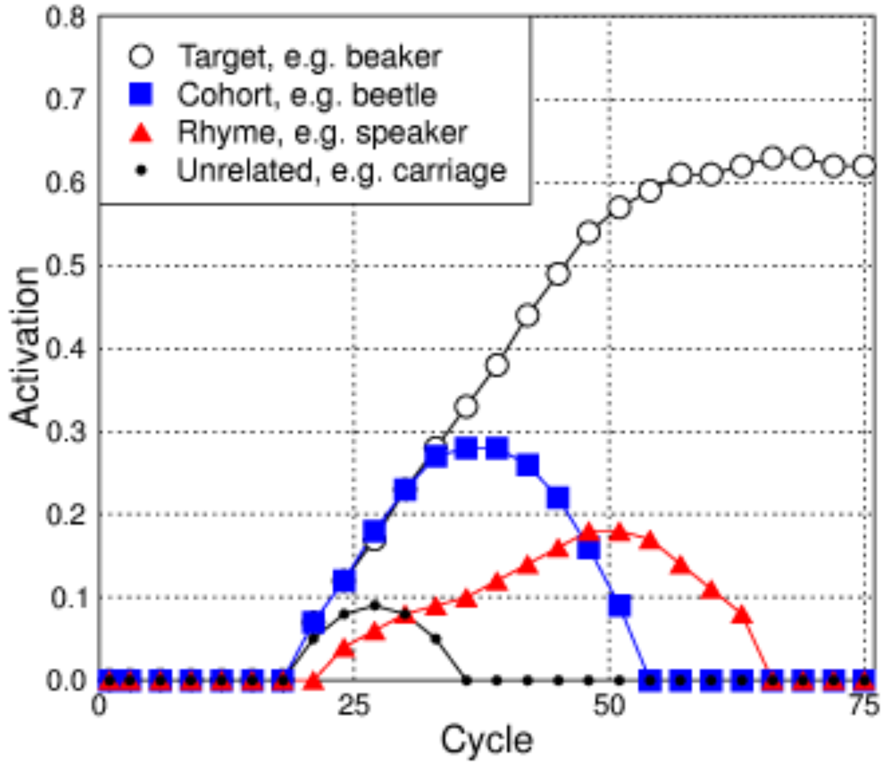
- Similar to interactive activation model **(交互激发模型)** but applied to speech recognition
- Connections between levels are bi-directional and excitatory
  → **top-down effects**
- Connections within levels are inhibitory producing **competition** between alternatives



(McClelland & Elman, 1986)

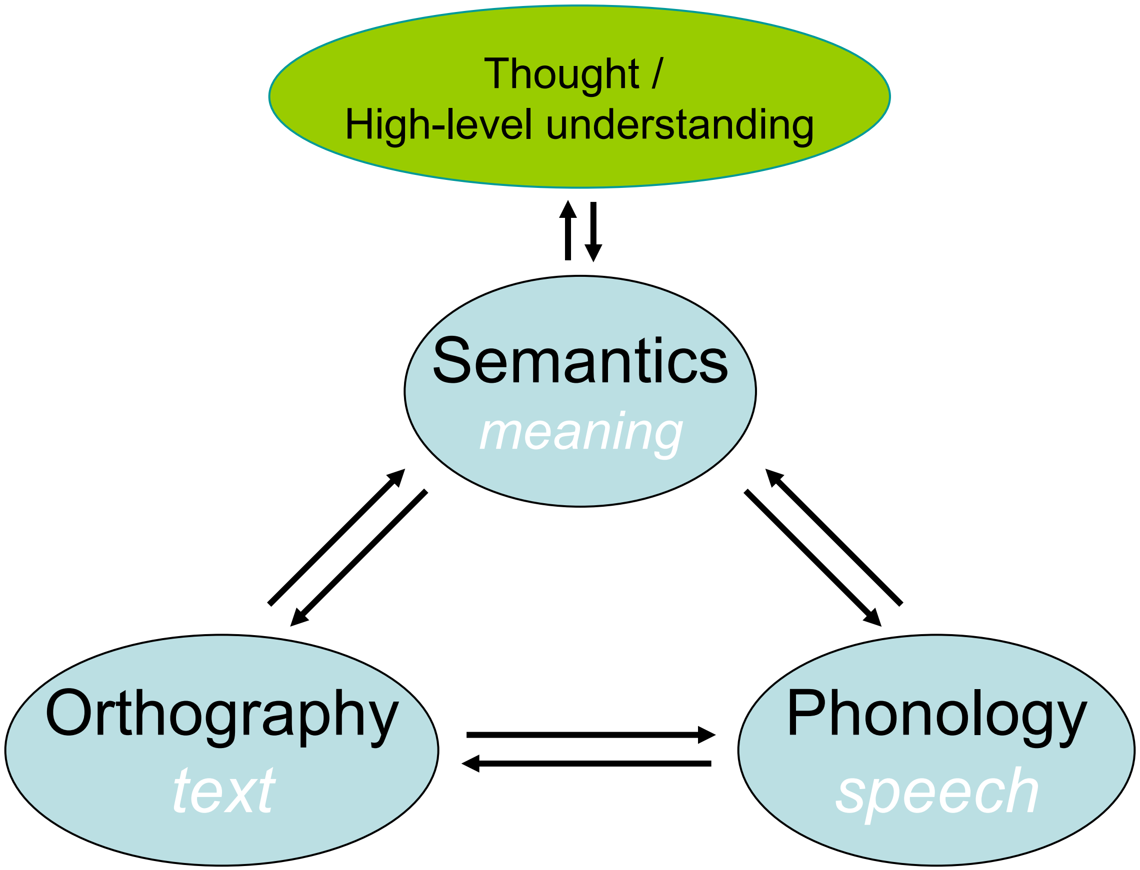
TRACE Model Predictions

- System activates candidate words that are consistent with current information
- Candidates compete with each other
- Winner is selected and competitors are inhibited



Semantic Representations and Naming Deficits

Triangle Model



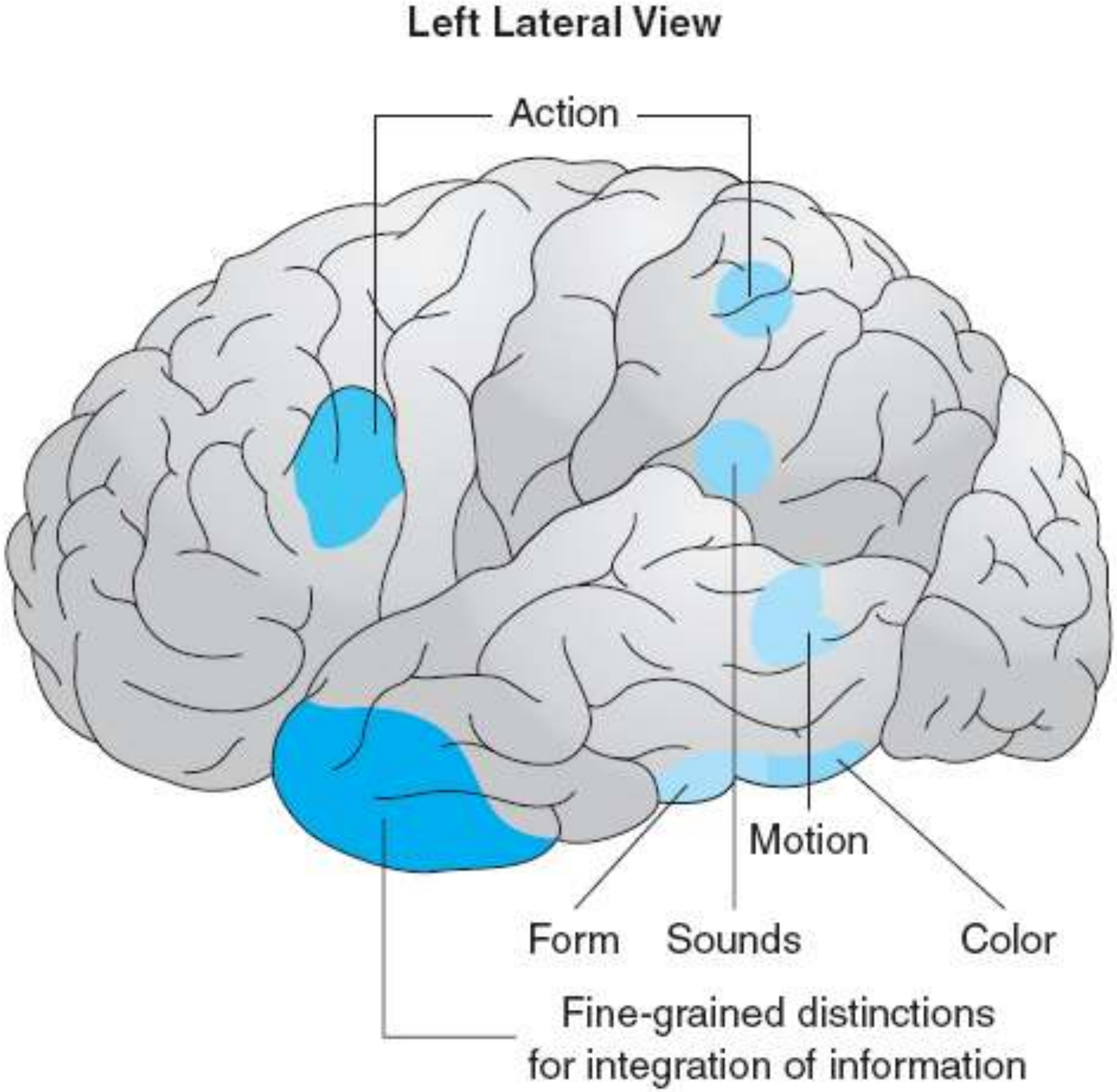
Connectionist framework for lexical processing, adapted from Seidenberg and McClelland (1989) and Plaut et al (1996).

Representing Meaning

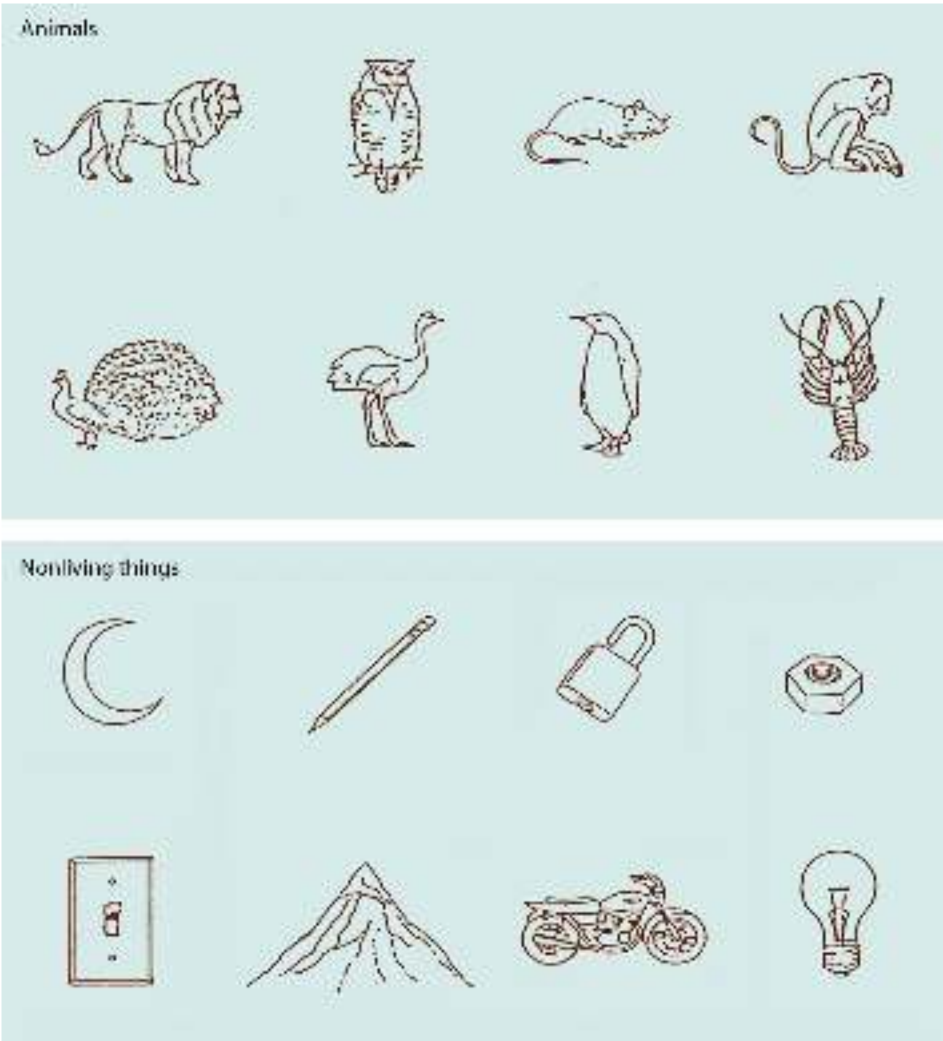
- No mental dictionary

- Instead, we appear to represent meaning with an interconnected (distributed) representation involving many different kinds of “features”

Representing Meaning



Category Specific Semantic Deficits



Evidence comes from patients with category-specific impairments (语义范畴特异性损伤)

Patients who have trouble naming living or non-living things

Warrington and Shallice (1984)

Definitions giving by patient JBR and SBY

Examples of definitions

JBR	Parrot: don't know	Tent: temporary outhouse, living home
	Daffodil: plant	Briefcase: small case used by students to carry papers
	Snail: an insect animal	Compass: tools for telling direction you are going
	Eel: not well	Torch: hand-held light
	Ostrich: unusual	Dustbin: bin for putting rubbish in
SBY	Duck: an animal	Wheelbarrow: object used by people to take material about
	Wasp: bird that flies	Towel: material used to dry people
	Crocus: rubbish material	Pram: used to carry people, with wheels and a thing to sit on
	Holly: what you drink	Submarine: ship that goes underneath the sea
	Spider: a person looking for things, he was a spider for his nation or country	Umbrella: object used to protect you from water that comes

Farah and McClelland (1991)

Explanation

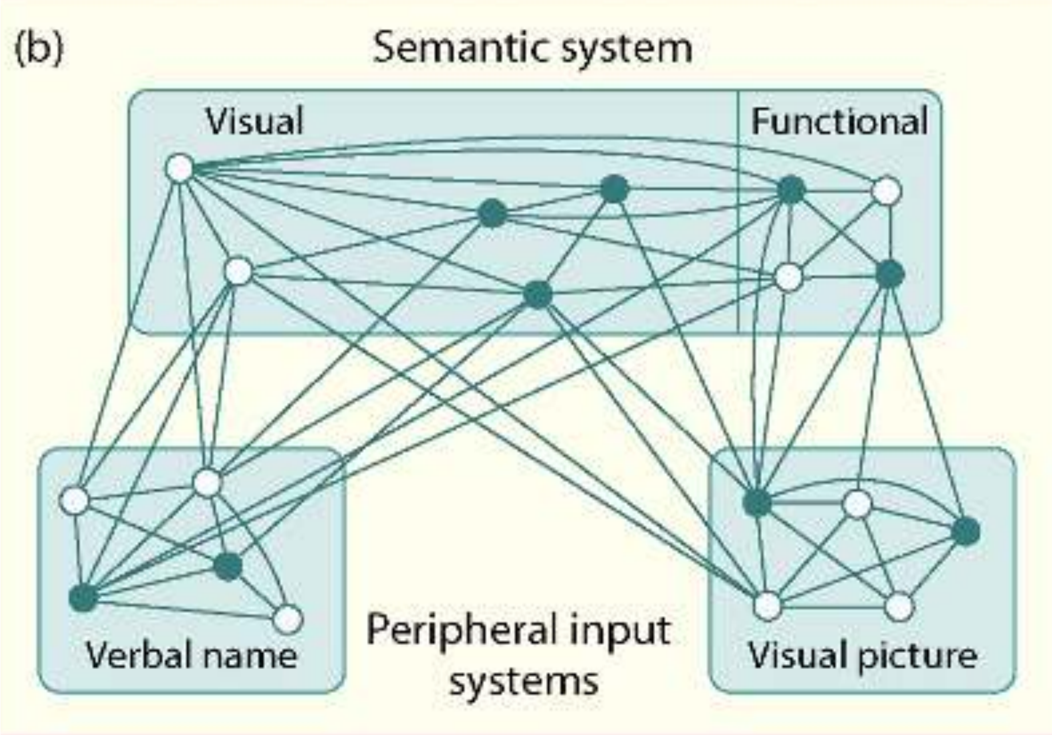
- One possibility is that there are two separate systems for living and non-living things
- More likely explanation: Living things are primarily distinguished by their **perceptual (知觉的)** attributes. Non-living things are distinguished by their **functional (功能的)** attributes → **Sensory functional approach**
  - Objects have visual and functional property features
  - Living things were represented by more visual than functional representational nodes

Sensory-Functional Approach

- Support for sensory-functional approach comes from a number of studies
- Example: Dictionary studies measured the ratio of visual to functional features:
  - **7.7:1** for living things
  - **1.4:1** for nonliving things

Farah and McClelland (1991)

A neural network model of category-specific impairments



A **single system** with functional and visual features

Model was trained to **associate** visual picture with the name of object using a distributed internal semantic representation

Farah and McClelland (1991)

A neural network model of category-specific impairments

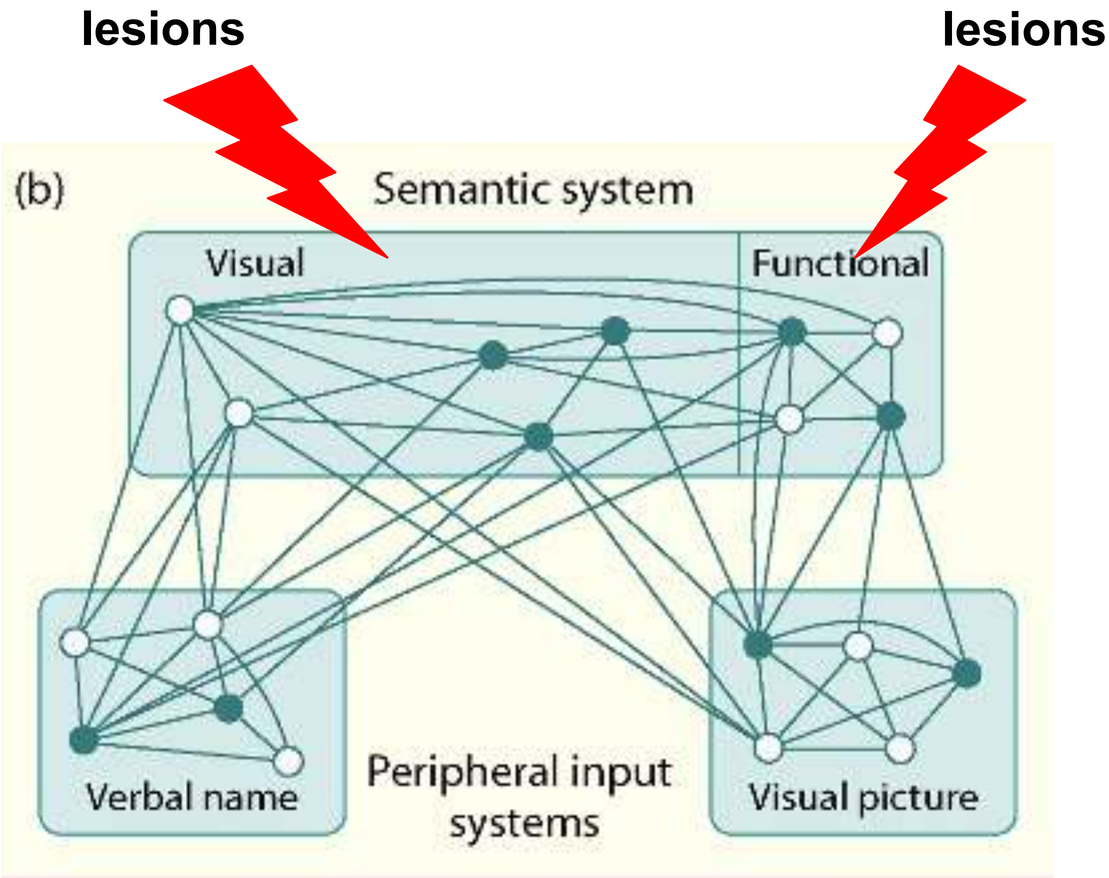
- A **single system** with functional and visual features. Model was trained to discriminate N living and nonliving things

- Two main layers: **semantic** and **input/output layer**. The two layers have bidirectional connections

- Network was trained to produce the appropriate verbal name output when presented with a picture. Also, the network was trained to produce the appropriate visual attributes when presented with a name as input. The weights between all nodes were tuned such that the network made few mistakes in this verbal and visual naming task

Farah and McClelland (1991)

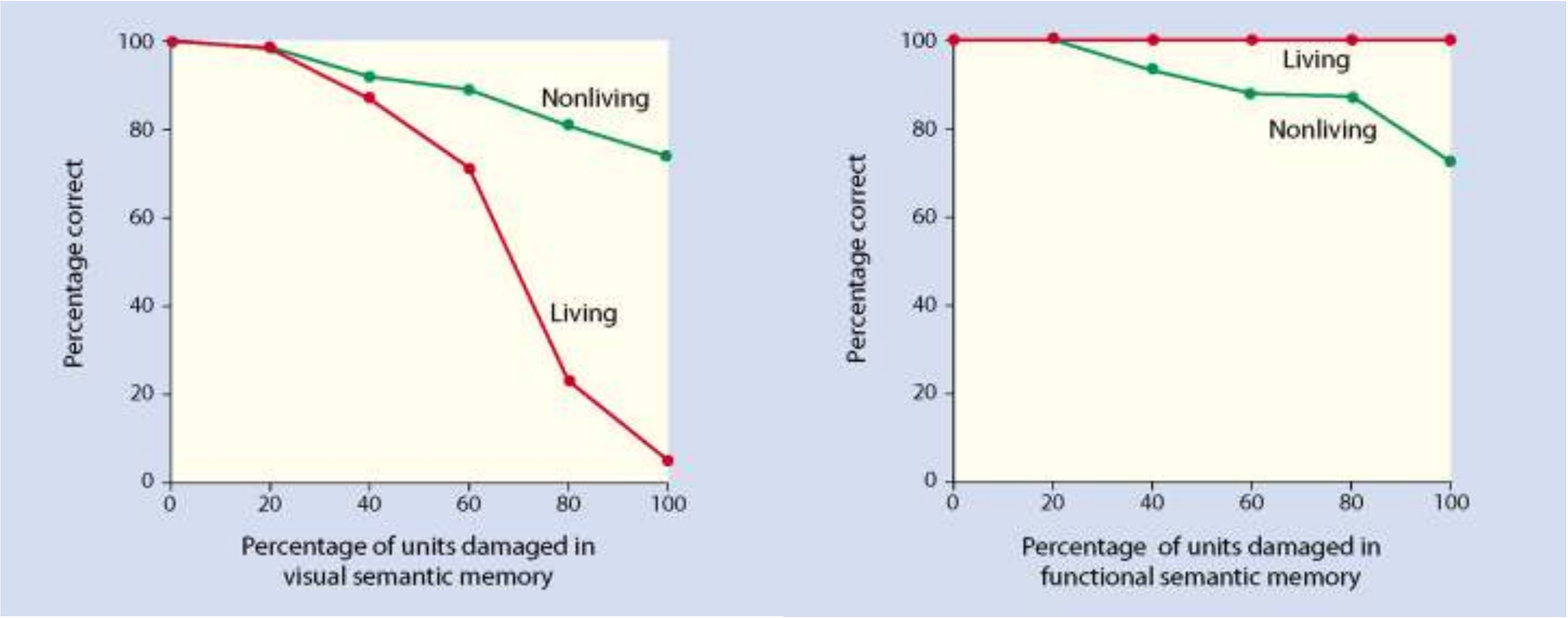
A neural network model of category-specific impairments



Simulate the effect of brain lesions

Farah and McClelland (1991)

Simulating the Effects of Brain Damage by "lesioning" the model



Visual Lesions: selective impairment of living things

Functional Lesions: selective impairment of non-living things

Farah and McClelland (1991)