- Meng, Y., Wynne, H., & Lahiri, A. (2021). Representation of “T3 sandhi” in mandarin: significance of context. *Language, Cognition and Neuroscience*, *36*(6), 791–808
  - Contextual phonological alternation
    - segmental assimilation
    - segmental dissimilation
    - tonal assimilation
    - tonal dissimilation
  - The role of lexical tone in spoken word recognition

Main views

- The “canonical representation” view: all T3 sandhi words are represented in the citation form (i.e. T3-T3), despite the tone changing on the surface.
- The “surface representation” view: T3 tone sandhi words are always represented in their surface form, that is, T3-T3 > T[2]-T3.
- The “abstract representation” view: the initial morphemes involved in the T3 sandhi have two surface forms, i.e. both T2 and T3, causing the lexical representation to be “abstract” and thus compatible to both surface forms.

Research question

- The views on the representation of T3 and particularly how T3 sandhi words are represented in the mental lexicon remain controversial.
- To further investigate the influence of context on T3 sandhi word processing.
- A series of cross-modal priming (or semantic) experiments.

– Exp 1

Table 2. Example of primes and targets in Experiment 1.

Auditory prime	Example	Visual target
T2	*fan2* "annoying"	反射
T3	*fan3* "opposite"	*fan3she4*
T4 (control)	*fan4* "meal"	"reflection"

Table 3. Mean Response Times (in ms) for Experiment 1.

Condition	Mean	SD	Priming effect (compared to T4)
T2	582	102	4
T3	565	102	−13**
T4 (control)	578	100	N/A

Note. ** $p < .001$. The negative value of priming effect represents faster response.

– This finding suggests that an isolated T2 syllable cannot prime a target beginning with T3 in a non-sandhi context.

– Exp 2

Table 4. Example of primes and targets in Experiment 2.

Auditory prime	Example	Visual target
semantically related T3	*nao3* "brain"	头部
T2	*nao2* "to scratch"	*tou2 bu4*
T4 (control)	*nao4* "noisy"	"head"

Table 5. Mean Response Times (in ms) for Experiment 2.

Condition	Mean	SD	Priming effect (compared to T4)
T2	584	99	−3
T3	573	106	−14**
T4 (control)	587	106	N/A

Note. ** $p < .01$. The negative value of priming effect represents faster response.

– the surface T2 primes did not activate targets that were semantically related to the T3 primes. In contrast, the priming effect by T3 primes which were semantically related to the targets was substantial and straightforward.

– Exp 3a & b

Table 6. Example of task design in Experiment 3. The "mediator" condition is never presented but the first morpheme is homophonous to the T3 condition and it is semantically related to the target.

Auditory Prime	Example	Mediator	Visual Target
T3	*da3* "to beat"	打扫 *da3sao3* "to clean"	清理 *qing1li3* "to clean up"
T2	*da2* "to answer"		
T4 (control)	*da4* "big"		

Table 7. Mean Response Times (in ms) for the prime-mediator word pairs in Experiment 3a.

Condition	Mean	SD	Priming effect (compared to T4)
T2	545	103	−22*
T3	542	98	−25***
T4 (control)	567	108	N/A

Note. * $p < .05$, *** $p < .001$. The negative value of priming effect represents slower response.

Table 8. Mean Response Times (in ms) for the prime-target word pairs in Experiment 3b.

Condition	Mean	SD	Priming effect (compared to T4)
T2	522	96	−15*
T3	521	98	−16*
T4 (control)	537	104	N/A

Note. *$p < .05$. The negative value of priming effect represents faster response.

- For instance, while nao2 “to scratch” is not by itself synonymous with tou2bu4 “head”, it could be if combined with other contextually-related morphemes. If this contextually-driven relationship is not strong enough to facilitate the processing of targets, transfer from T2 to T3 will be blocked.
- the T2 primes facilitated the processing of targets (when compared with T4 controls), suggesting that they could activate the mediating T3 tone sandhi words only in the correct sandhi context.

- The “surface representation” view postulates that T3 sandhi words are represented as the surface form T[2]-T3, which ought to allow T2 primes to facilitate lexical decision while T3 primes should not. Our results do not support this.
- The phonological interference applies that input T2 activates the underlying representation of T3 only in the context of another following T3, and the re-writing rule is justified by context.
- The re-writing is only permitted when the following syllable immediately follows is also a T3.

Week 6: Word recognition

- In spoken languages, sounds that have no intrinsic meaning are combined to form larger, meaningful units like morphemes and words. But in logographic writing systems, the smallest units of combination do have intrinsic meaning.

- What happens when a reader who's developed an efficient network for reading Chinese learns to read English, or vice versa?
- Does a person's first language reading network become recruited for the task of reading the second language?

- A study led by Jessica Nelson (2009) suggests that it depends on which writing system you start out with. Evidence from fMRI studies showed that Chinese subjects who learned English as a second language were able to use the same reading networks for both English and Chinese.
  - But the reverse wasn't true; English speakers who learned to read Chinese showed different patterns of brain activity for the two languages.
  - The authors suggest that it may be possible to read English as if it were Chinese—that is, by recognizing whole words and matching them to word meanings, rather than by sounding out the individual phonemes of a word.
  - However, since using a "sounding-out" strategy for reading Chinese is not really viable, English learners of Chinese would have been forced

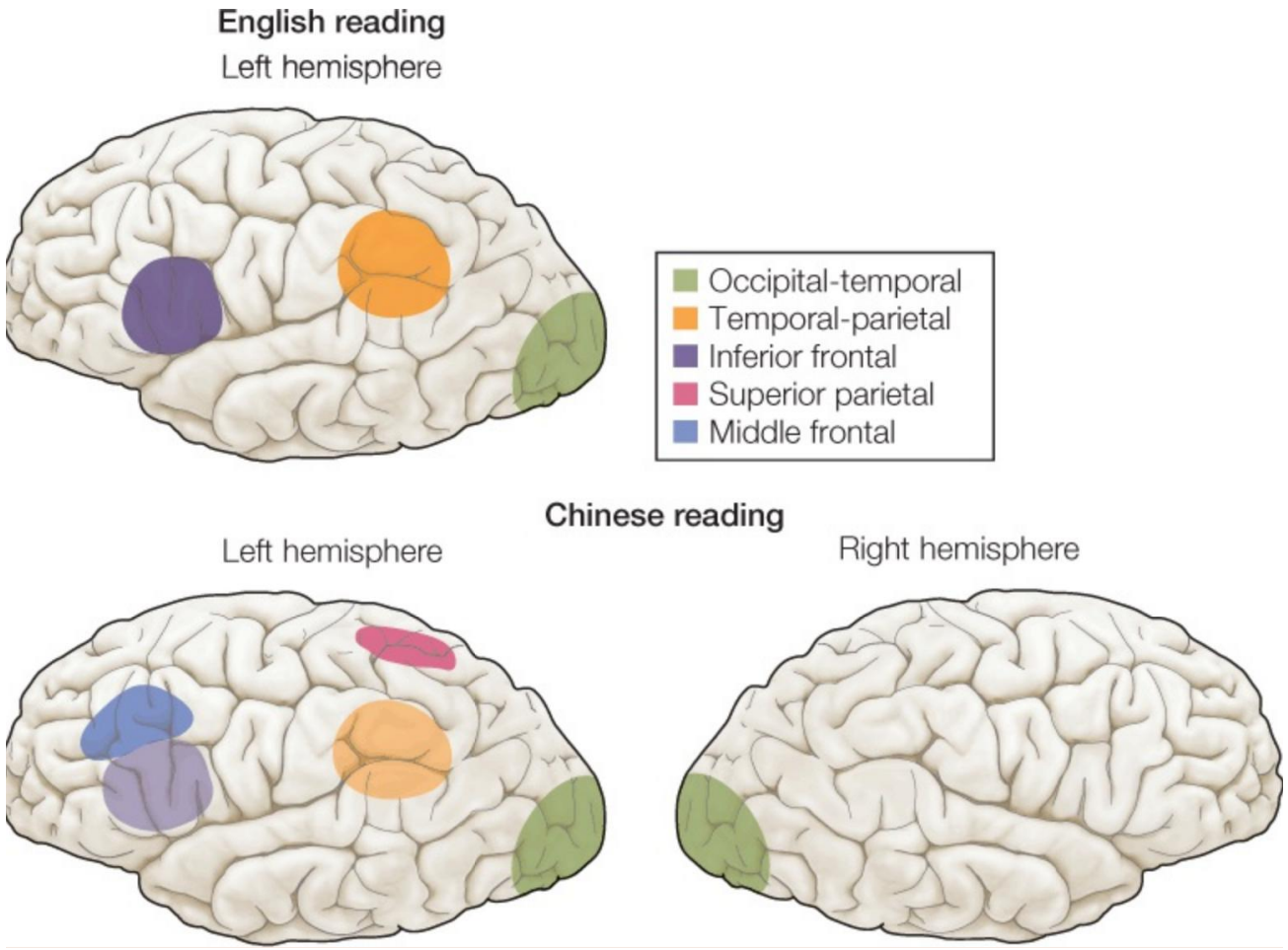
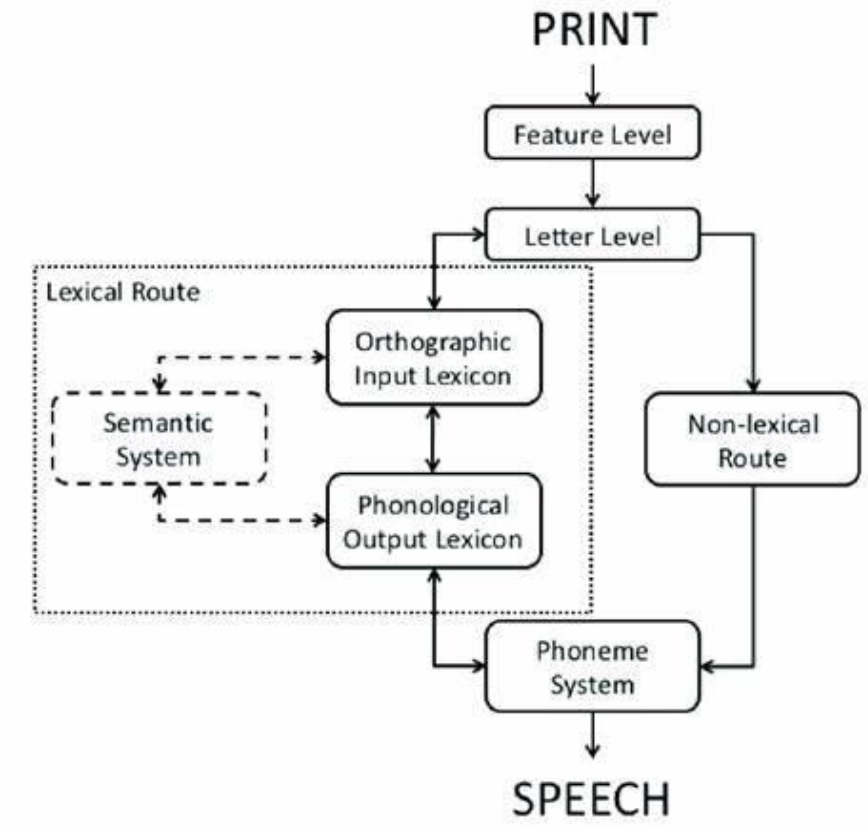


Figure 8.12 A rough diagram of English and Chinese reading networks, as identified by Perfetti et al. (2010). Chinese readers show more bilateral activity in occipital-temporal regions (green) and activity in the left middle frontal gyrus (blue). Chinese reading involves reduced activity in inferior frontal areas (purple) and in temporal-parietal regions (gold). (Adapted from Perfetti et al., 2010, in Cornelissen et al. (eds.), *The neural basis of reading*, p. 147.)

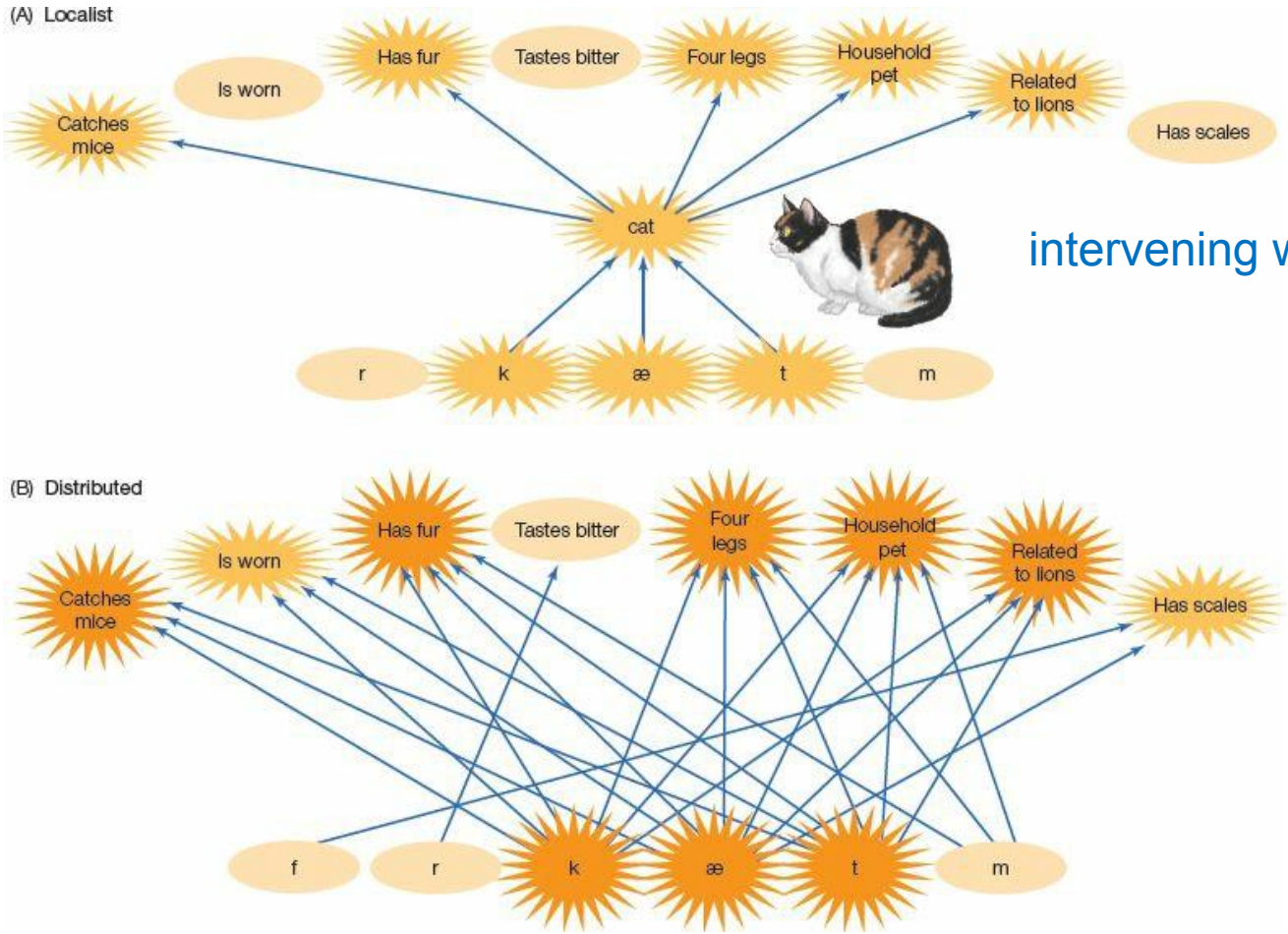
Dual route model

- the lexical route is most effective with skilled readers who can recognize words that they already know.
- The non-lexical is a mechanism for decoding novel words using existing grapheme-to-phoneme rules in a language. This mechanism operates through the identification of a word’s constituent parts (such as graphemes) and applying linguistic rules to decoding



- English
  - lexical route: apple, dog
  - non-lexical route: Pneumonoultramicroscopicsilicovolcanoconiosis
- Chinese
  - lexical route: 我,
  - non-lexical route 

- We've assumed that word representations themselves are discrete units.
  - individual nodes that become activated or inhibited as entire units.
  - discrete units (or containers) vs. bundles of features



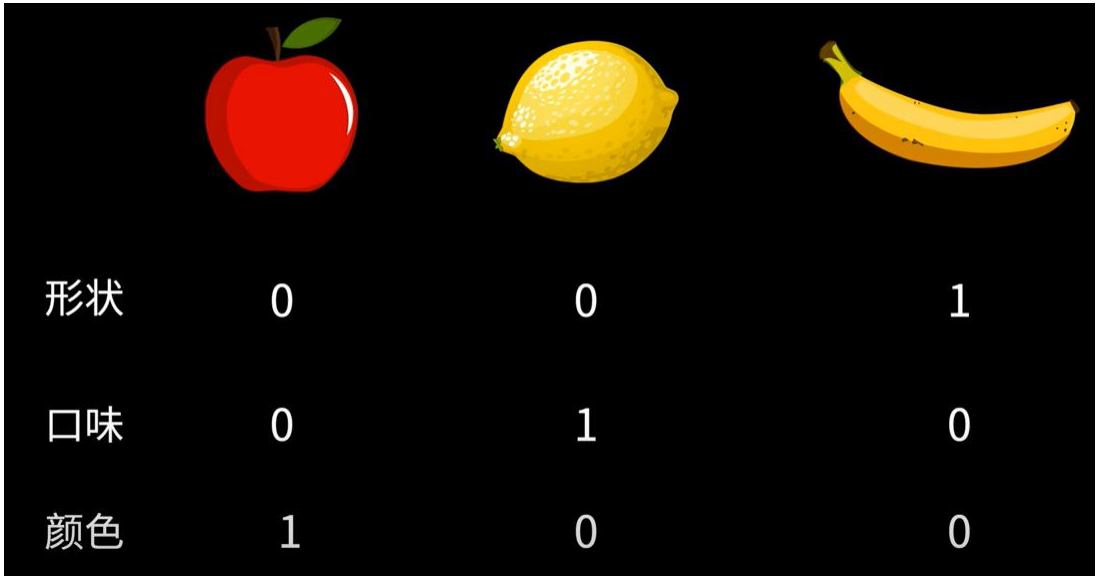
intervening word nodes

Local (局部) vs. Distributed (分布)word representations

- Local representation:
  - The existence of orthographic lexicon
  - *“if the visually-presented stimulus is represented in the orthographic lexicon, then it is a word”*

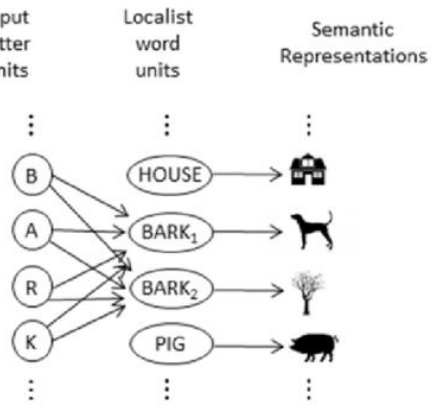
--- Rastle, 2007

- Distributed representation (Harm & Seidenberg, 2004; Plaut, 1997;
  - Deny the existence of lexicon
  - The information about known words is coded in a distributed manner as learned patterns of activation over a large body of units
  - There are no individual units for known words



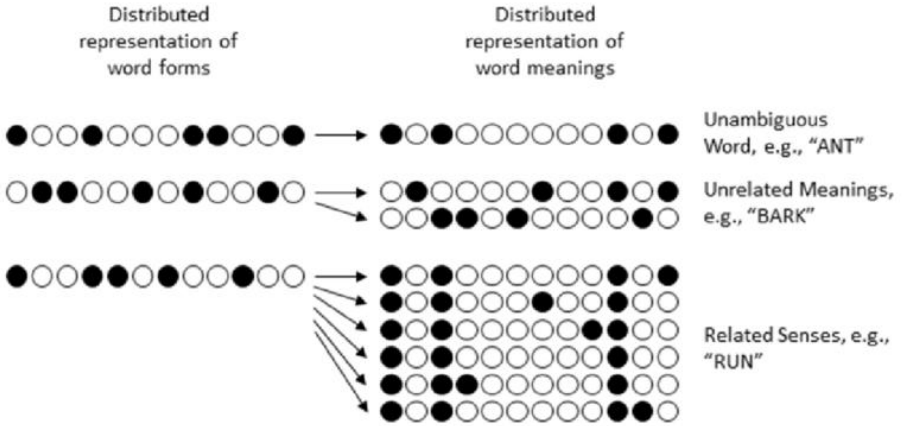
- **localist approach** assume that each word that we know is represented by a single entry inour mental lexicon. These entries correspond to individual ‘nodes’ or ‘units’

(a) Localist Representations of Ambiguous words

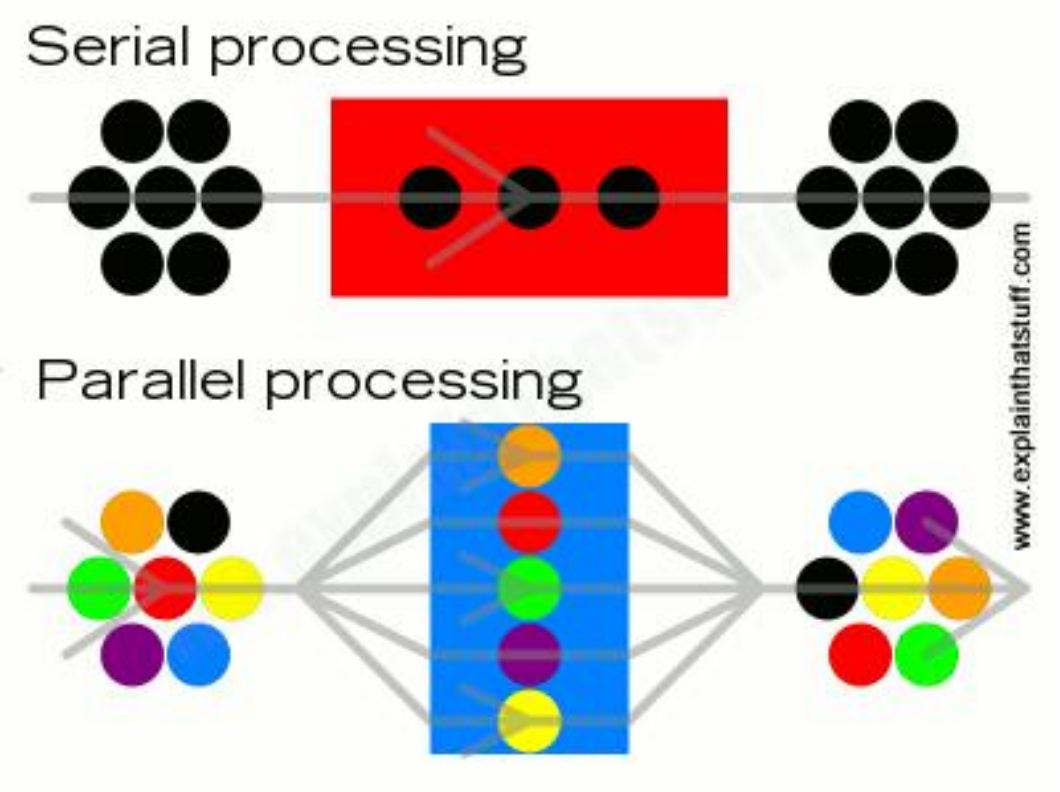


- **distributed approach** assumes that each word that we know is represented as a unique pattern of activation across sets of units that collectively represent its form (i.e.spelling/orthography or sound/phonology) and its meaning

(b) Distributed Representations of Ambiguous words

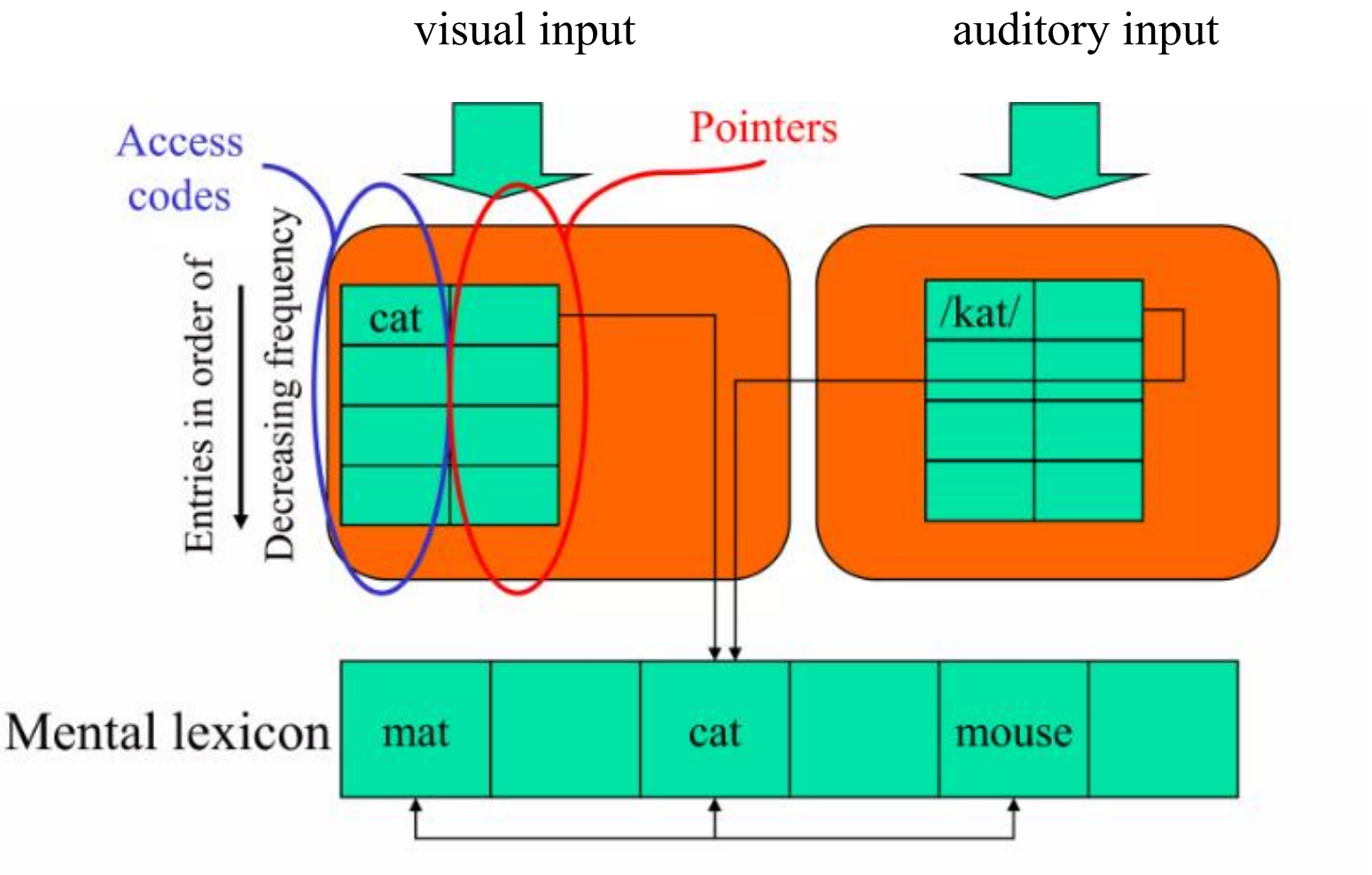


- Serial processing along the line of text (Reichle, Liversedge, Pollatsek, & Rayner, 2009)
- Others argue for a graded allocation of attention and parallel processing of multiple words (Engbert, Nuthmann, Richter, & Kliegl, 2005; Kennedy, 2000)



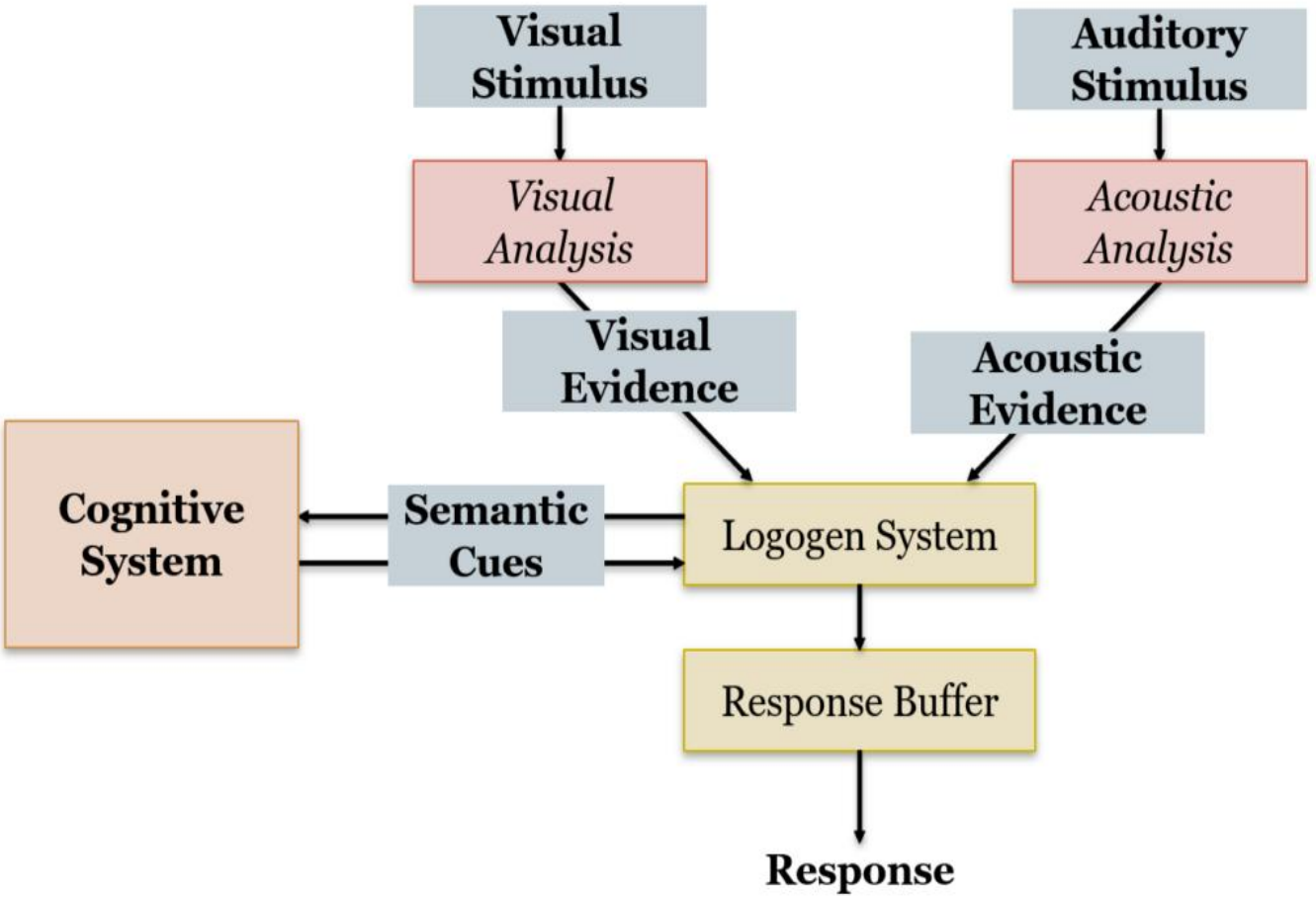
- Two classes of models
  - Models believing in local word representation
    - Search Models (Forster, 1976; Forster & Murray, 2004; Rubenstein et al, 1971)
    - Logogen model (Morton, 1969)
  - Models believing in distributed word representation
    - Connectionist models
    - Parallel distributed processing model (PDP model) (Plaut, 1996)

Search model (Forster, 1976)



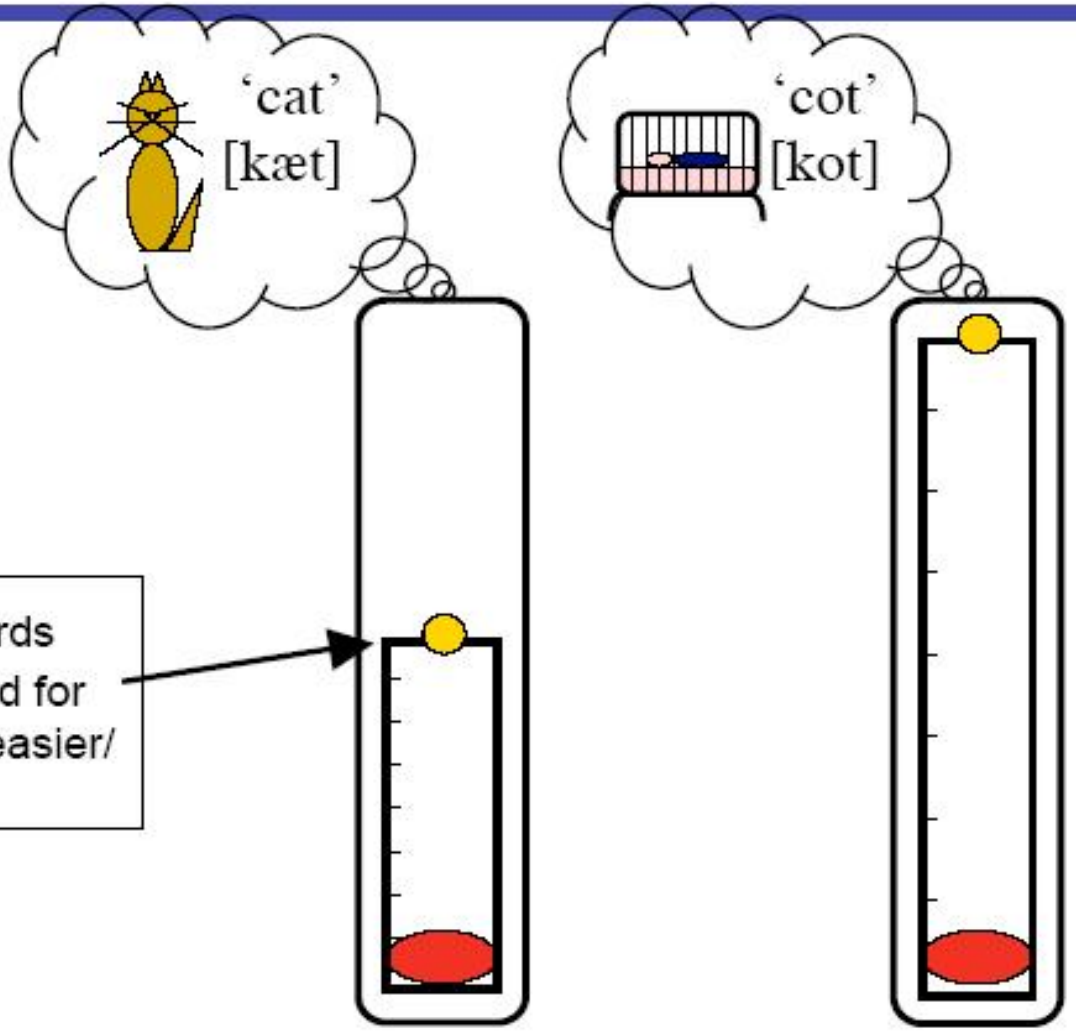
- a complete perceptual representation of the perceived stimulus is constructed
  - three access files:
    - orthographic (reading)
    - phonological (listening)
    - syntactic/semantic (production)
- Access files are organized in a series of bins
- Comparison of the representations of the perceived stimulus are one by one against the representations in the bins
- serial position within a bin affects the speed of access (ordered by word frequency, temporary reordering of bins in response to recent encounter)
- meanings are not stored in the access files, but a master file in semantic memory indexed by the pointer

Logogen model (Morton, 1969)

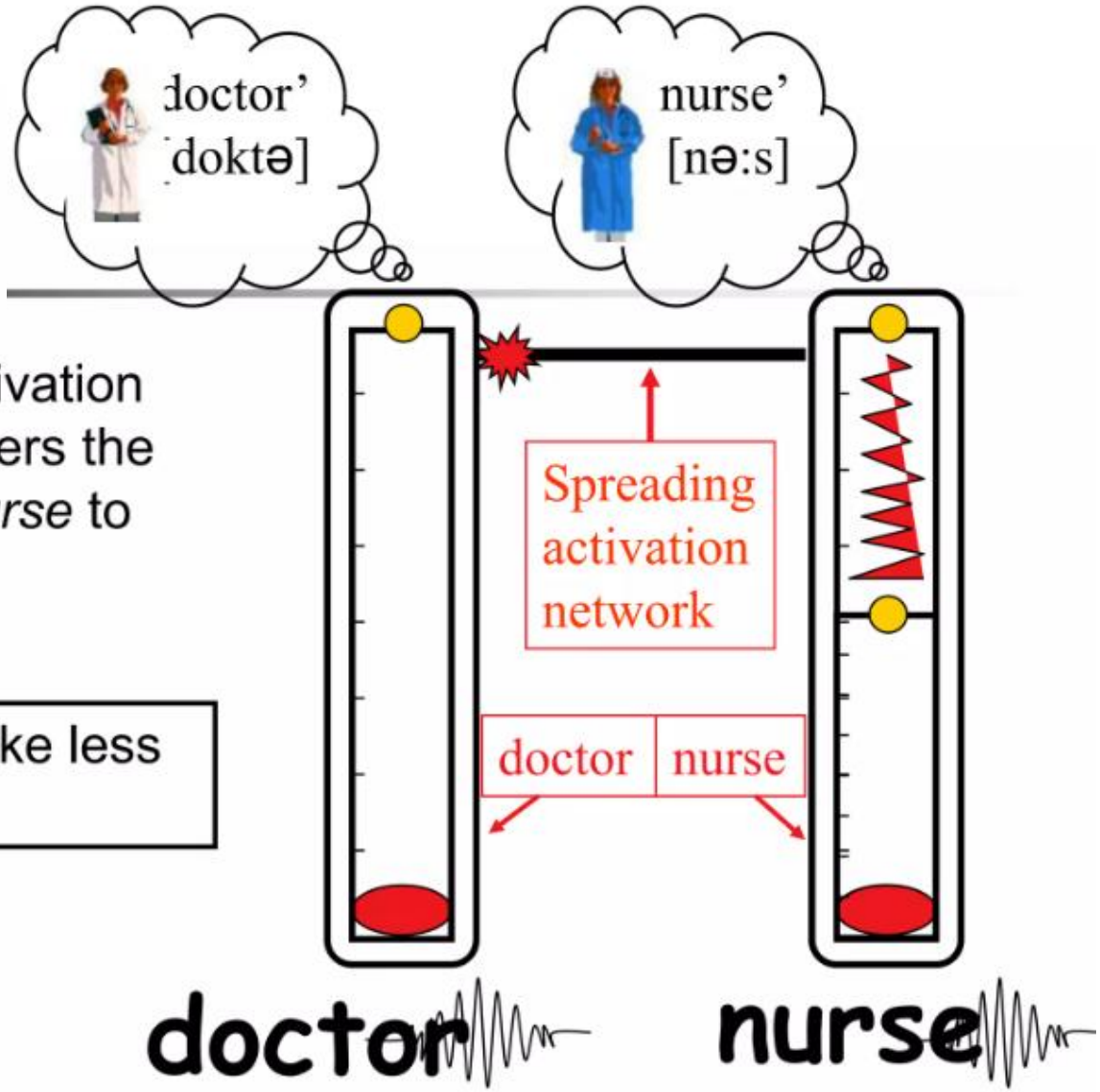


- According to logogen model, perceptual input feeds into feature counter called logogen (特征计数器)
- Each word is represented by a logogen. If the perceptual input contains a feature of a particular word, then the feature count of the logogen increases
  - actiaved by sensory input
  - by contextual information
- each logogen has a threshold at which it fires. When the word fires, the word is identified and the meaning is avaliable.

Frequency effect in Logogen



— High frequency words have a lower threshold for firing - they are fired easier/ faster



- Spreading activation from *doctor* lowers the threshold for *nurse* to fire

  – So *nurse* take less time to fire

- Connectionist model (Gaskell et al., 1997)

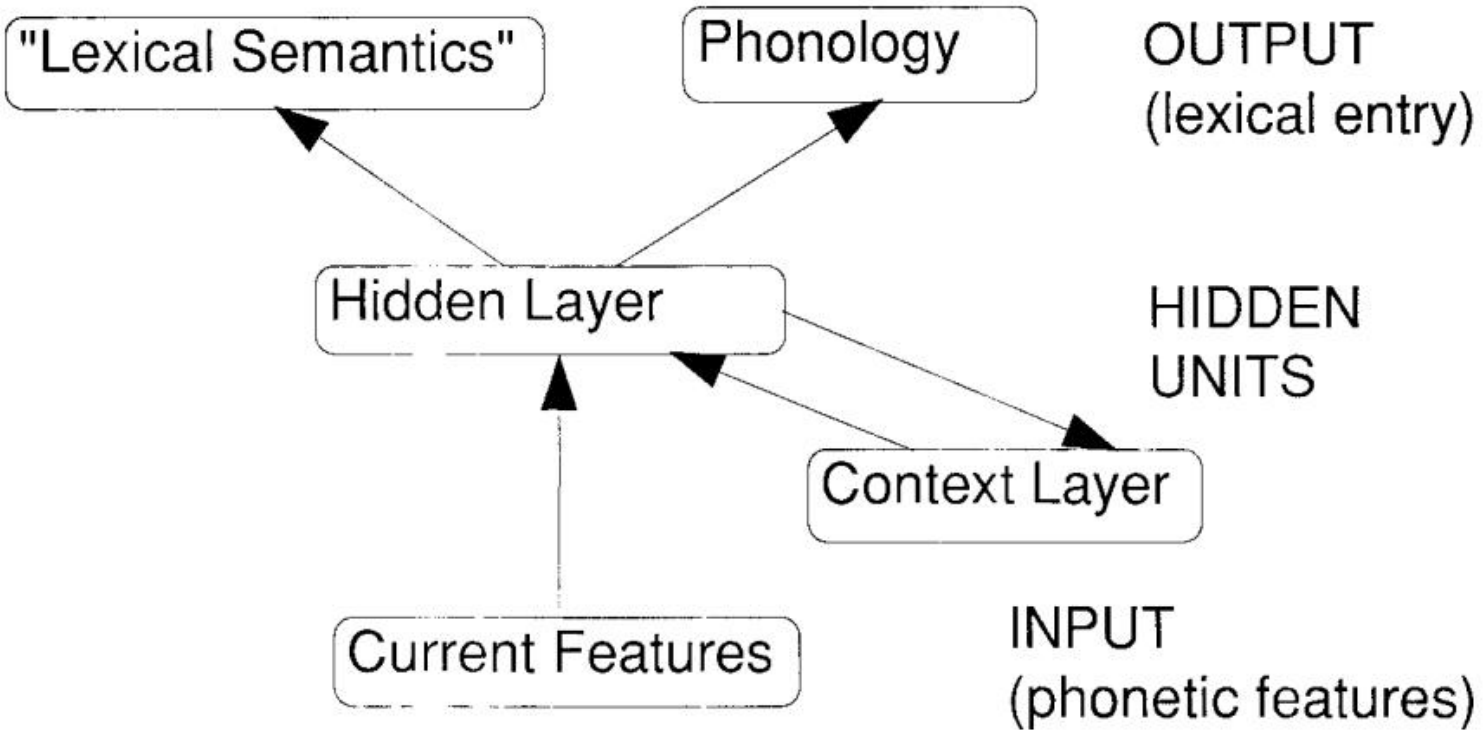


FIG. 1. A distributed model of speech perception.

- Parallel, distributed
- weighted connections between elements
- intermediate processors (now known as "hidden layers") alongside input and output units, and used a sigmoid activation function instead of the old "all-or-nothing" function

Neural network