Comprehensive Exam Presentation

Routes to Visual Word Recognition modelling : A comparative analysis of 3 computational models of reading aloud

Alexandre HERBAY

December 16th 2016

My Question

Discuss at least
two **current models** of **word recognition**,
considering the strengths and weaknesses of each.
Provide data to support your critical analysis.

Narrowing down

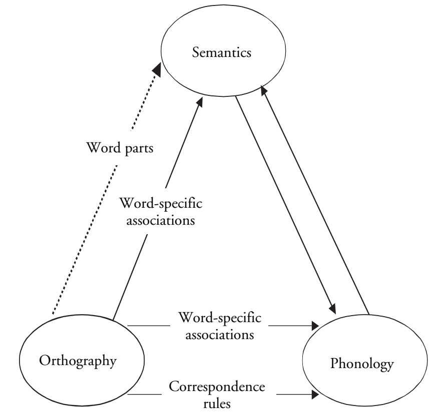
Spoken or **Visual Word recognition**

- Written words

> Theoretical models

Box-And-Arrow or **Computational Models**

- Computational = detailed algorithmic procedure that is computable and allows **detailed simulations**

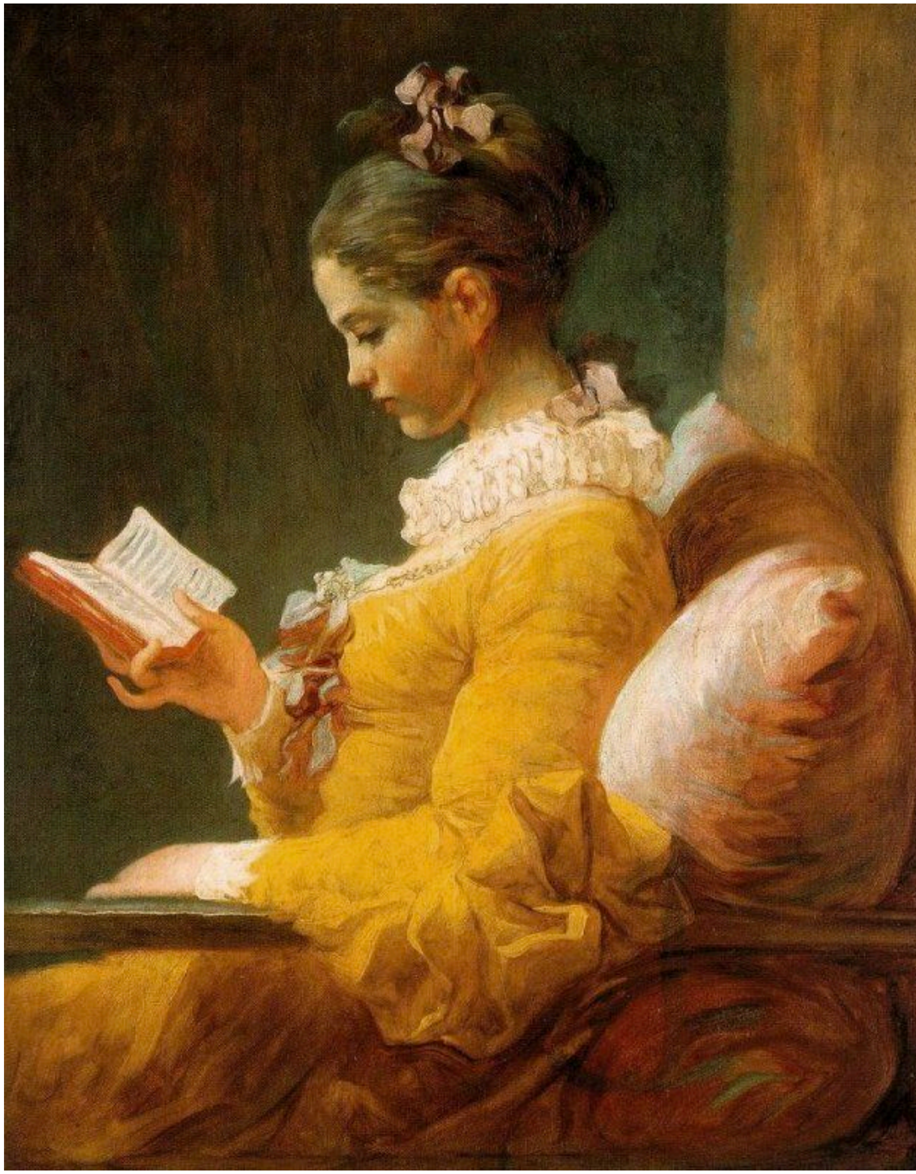


An architecture of the reading system (redrawn from Baron, 1977).

How many models ?

- **3 models**
  - 2 opposite approaches and their successor

Reading is quickly recognizing
a sequence of words
and assemble them in
a sentence to which we
attribute some meaning
in some context



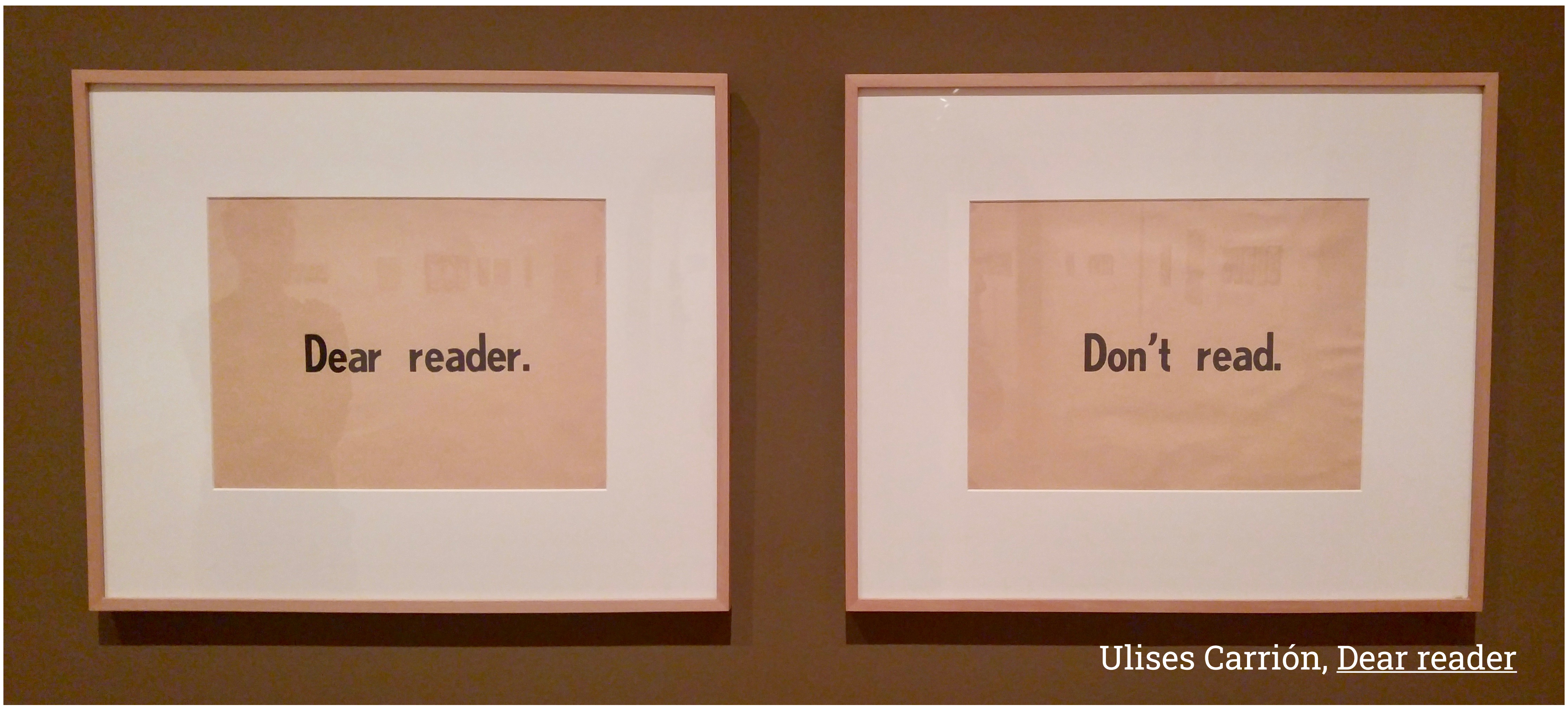
La liseuse, Fragonard (~1770)

Then reading is starting by recognizing words

An automatic process

stroop effect >

YELLOW
RED
GREEN



Ulises Carrión, Dear reader

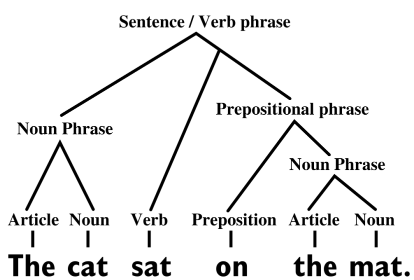
Words as reference level of reading

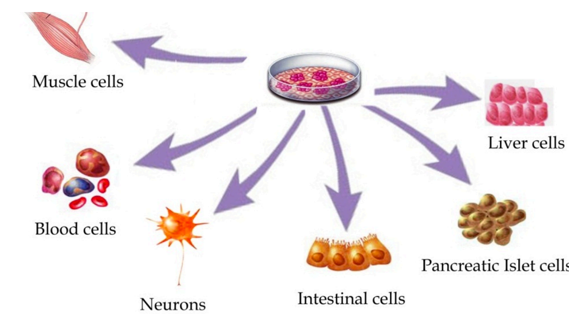
Words are like cells in biology or atoms in physics

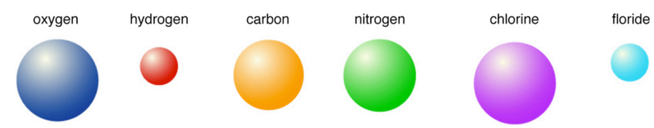
(Balota, 1994)

cat mat

sat the on







A word is a combination of three things

A VISUO-ORTHOGRAPHIC FORM

cat

A PHONOLOGICAL REPRESENTATION

/kat/

A MEANING

Reading a word = transforming a sequence of letters into a phonological representation or a meaning

A VISUO-ORTHOGRAPHIC FORM

cat

READING ALOUD → **A PHONOLOGICAL REPRESENTATION**

/kat/

READING SILENTLY → **A MEANING**

Two main tasks

- Lexical decision
- **Naming (reading aloud)**

Priming (masked or not) ;

Tracking eye-movement

Identification of degraded words ;

Category verification ;

Relatedness or rhyme judgments

…

Jeune fille lisant une lettre , Raout (1872)

A complete model would

explain the What, Where, When and How of Visual Word Recognition

Carreiras (2013)

A complete model would

explain the What, Where, When and How of Visual Word Recognition

WHAT

basic elements at play

HOW

the mechanisms to process them

WHERE

neural circuitry associated with this processing

WHEN

the time course of this processing

Carreiras (2013)

Different kind of models

Table 1. Major computational models of reading organised in terms of their primary focus[a,b]

Model	Style	Task	Phenomena	Large lexicon
Models of visual word recognition				
IA [11,22]	IA	PI	Word-superiority effect	
Multiple read-out [3]	IA	PI, LD	Word-superiority effect	
SCM [2]	IA	LD, MP	Letter order	
BR [4–6]	Math/comp	LD, MP	Word frequency, letter order, RT distribution	√
LTRS [8]	Math/comp	MP, PI	Letter order	
Overlap [66]	Math/comp	PI	Letter order	
Diffusion model [30]	Math/comp	LD	RT distribution, word frequency	
SERIOL [7]	Math/comp	LD, MP	Letter order	
Models of reading aloud				
CDP++ [13]	Localist/symbolic	RA	Reading aloud	√
DRC [12]	IA	RA, LD	Reading aloud	
Triangle [24,25]	Distributed connectionist	RA	Reading aloud	
Sequence encoder [15]	Distributed connectionist	RA	Reading aloud	√
Junction model [50]	Distributed connectionist	RA	Reading aloud	√
Models of eye-movement control in reading				
E-Z reader [17,18]	Symbolic	R	Eye movements	
SWIFT [19]	Symbolic	R	Eye movements	
Model of morphology				
Amorphous discriminative learning [16]	Symbolic network	Self-paced reading, LD	Morphology	√

[a]The table also indicates the modelling style or framework, the main task that the model simulates, the main phenomena that the model simulates (not exhaustive), and whether the model uses a realistically sized lexicon. Note that the review concentrates on 'Models of visual word recognition'.

[b]Abbreviations: Math/comp, mathematical or computational; LD, lexical decision; PI, perceptual identification; RA, reading aloud; MP, masked priming; R, natural reading.

perceptual ID
lexical decision

reading aloud

Norris (2013)

THREE MODELS

PARALLEL DISTRIBUTED PROCESSING (PDP)

DUAL-ROUTE CASCADED (DRC)

CONNECTIONNIST DUAL PROCESS + (CDP+)

PLAN

4. LIMITS AND CONCLUSION

MODELS ARCHITECTURES

1

THREE MODELS

Parallel Distributed Processing (PDP)

Dual-Route Cascaded (DRC)

Connectionnist Dual Process + (CDP+)

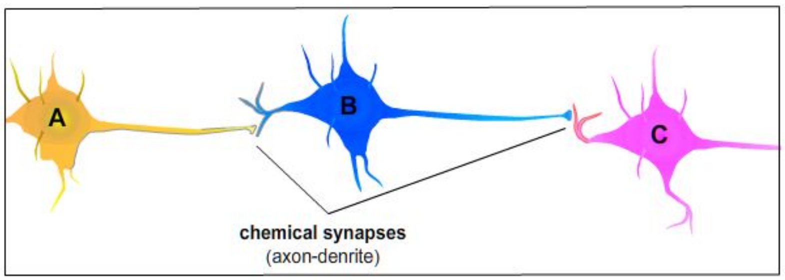
Parallel Distributed Processing Models (PDP)

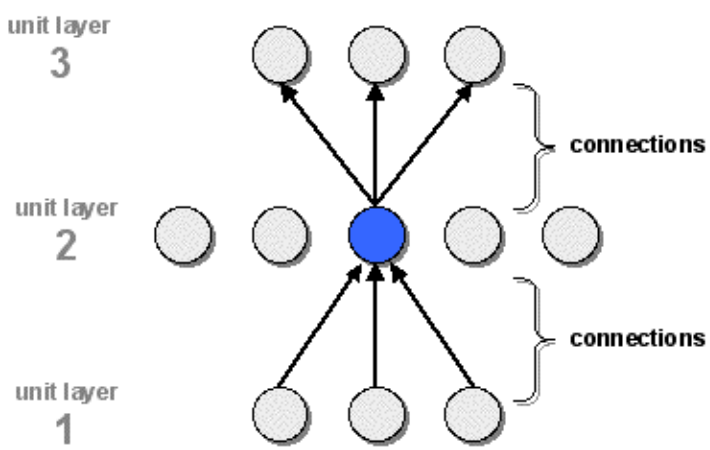
Connectionist model

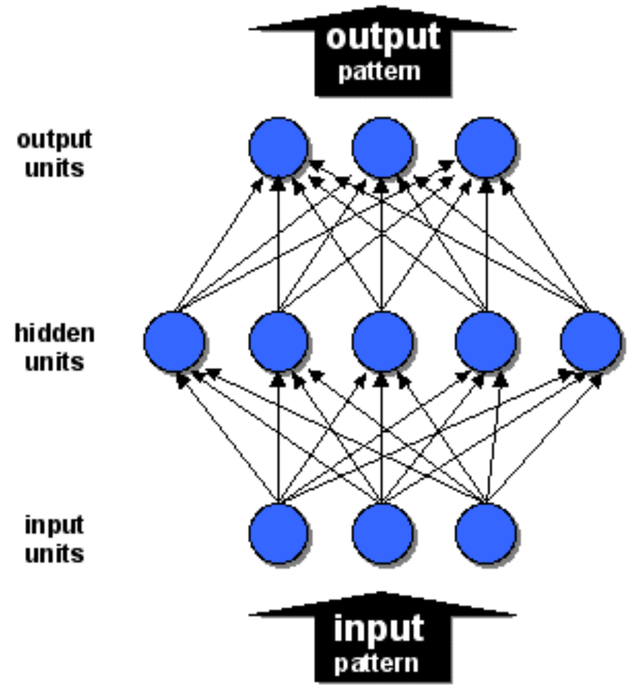
weighted connections

learning algorithm

hidden layers





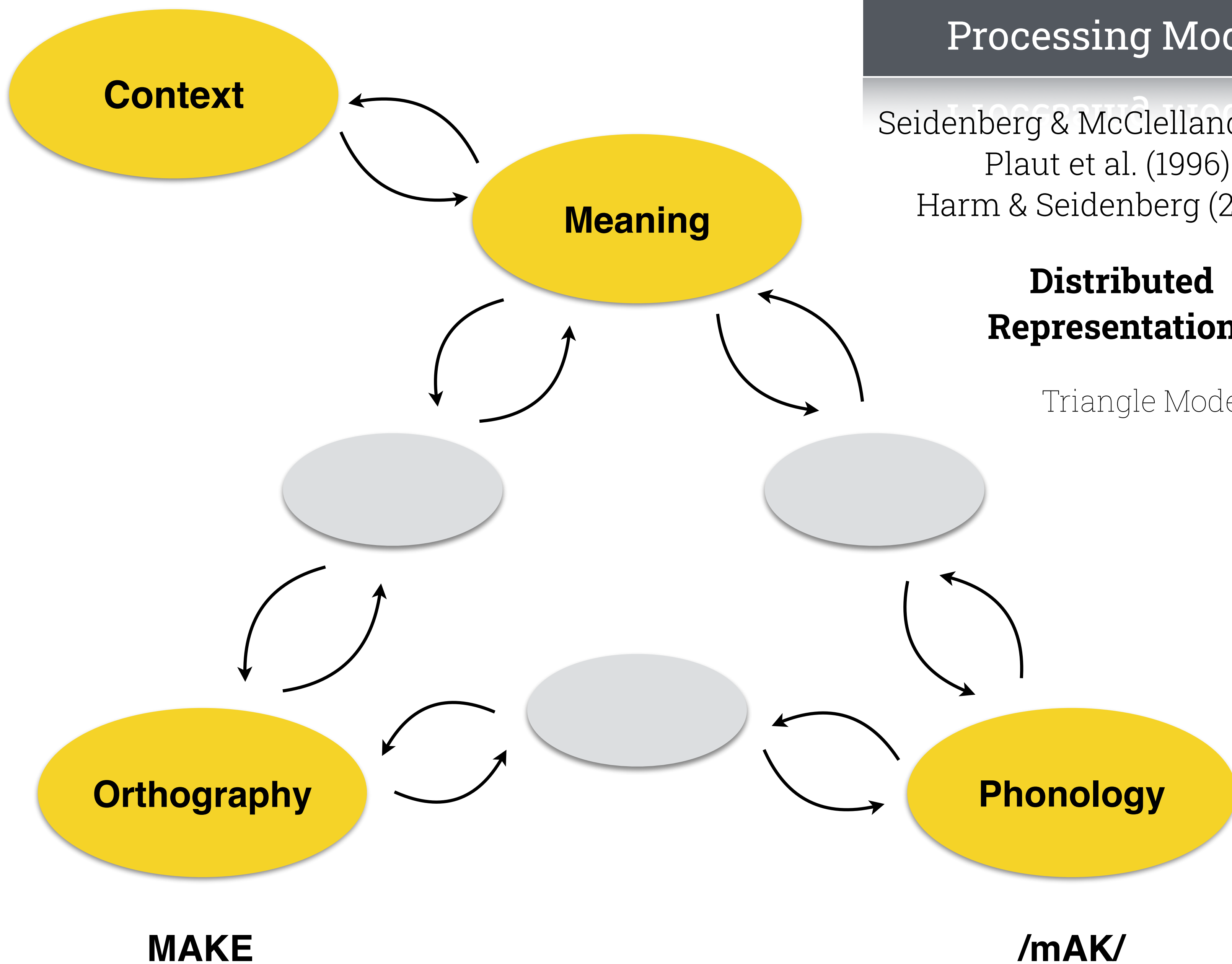


Paralled Distributed Processing Model

Seidenberg & McClelland (1989)
Plaut et al. (1996)
Harm & Seidenberg (2004)

Distributed Representations

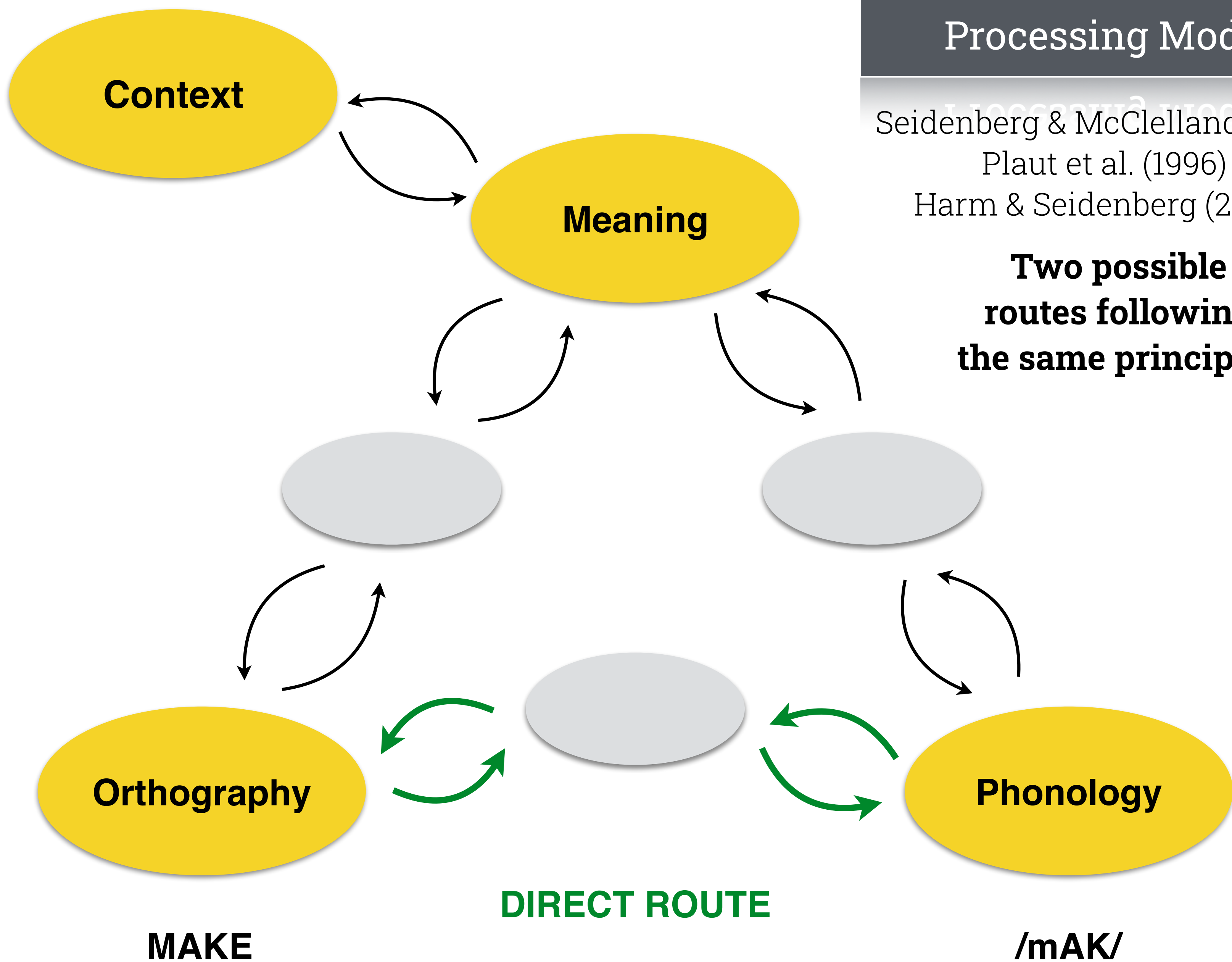
Triangle Model



Parallel Distributed Processing Model

Seidenberg & McClelland (1989)
Plaut et al. (1996)
Harm & Seidenberg (2004)

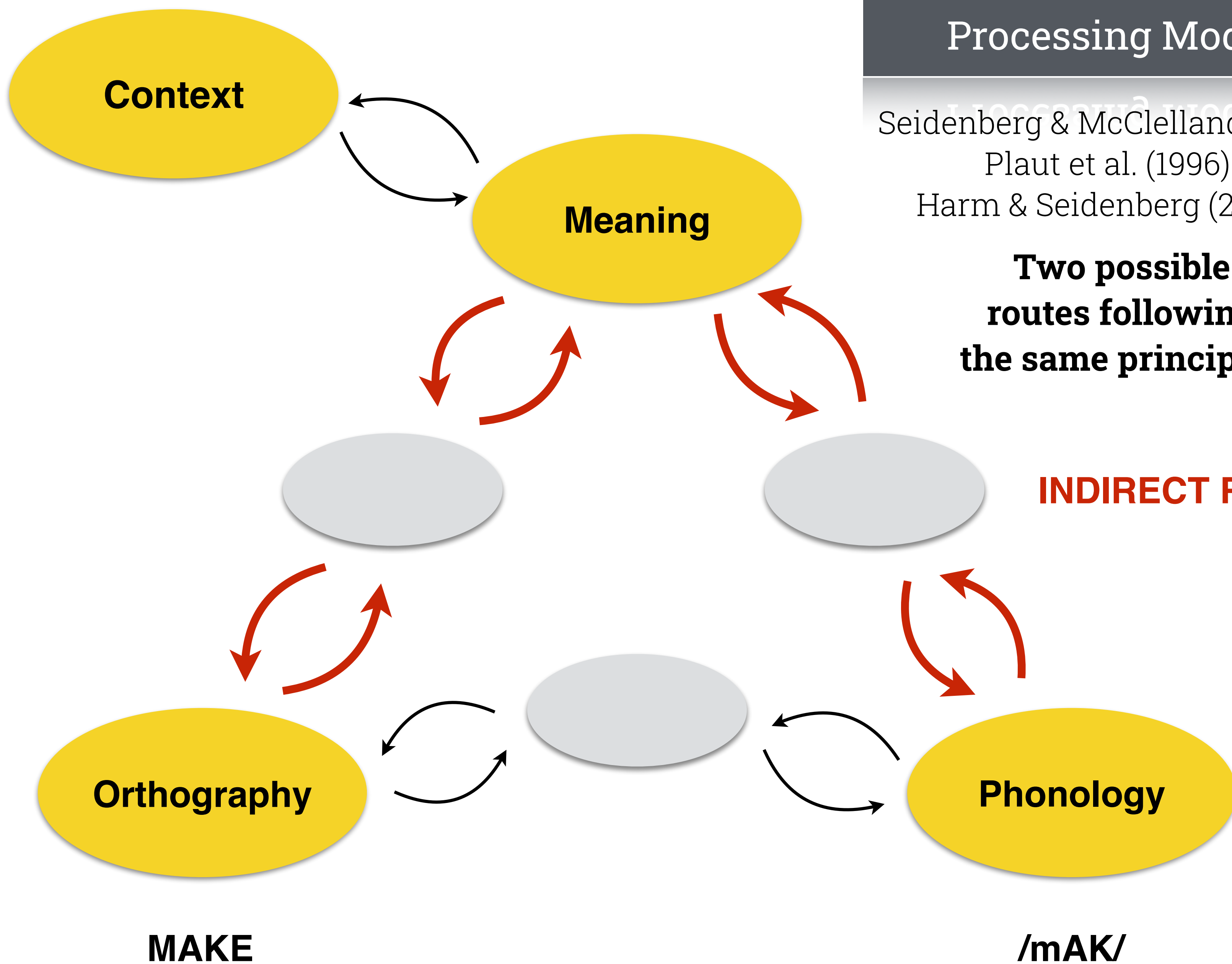
Two possible routes following the same principles



Parallel Distributed Processing Model

Seidenberg & McClelland (1989)
Plaut et al. (1996)
Harm & Seidenberg (2004)

Two possible routes following the same principles

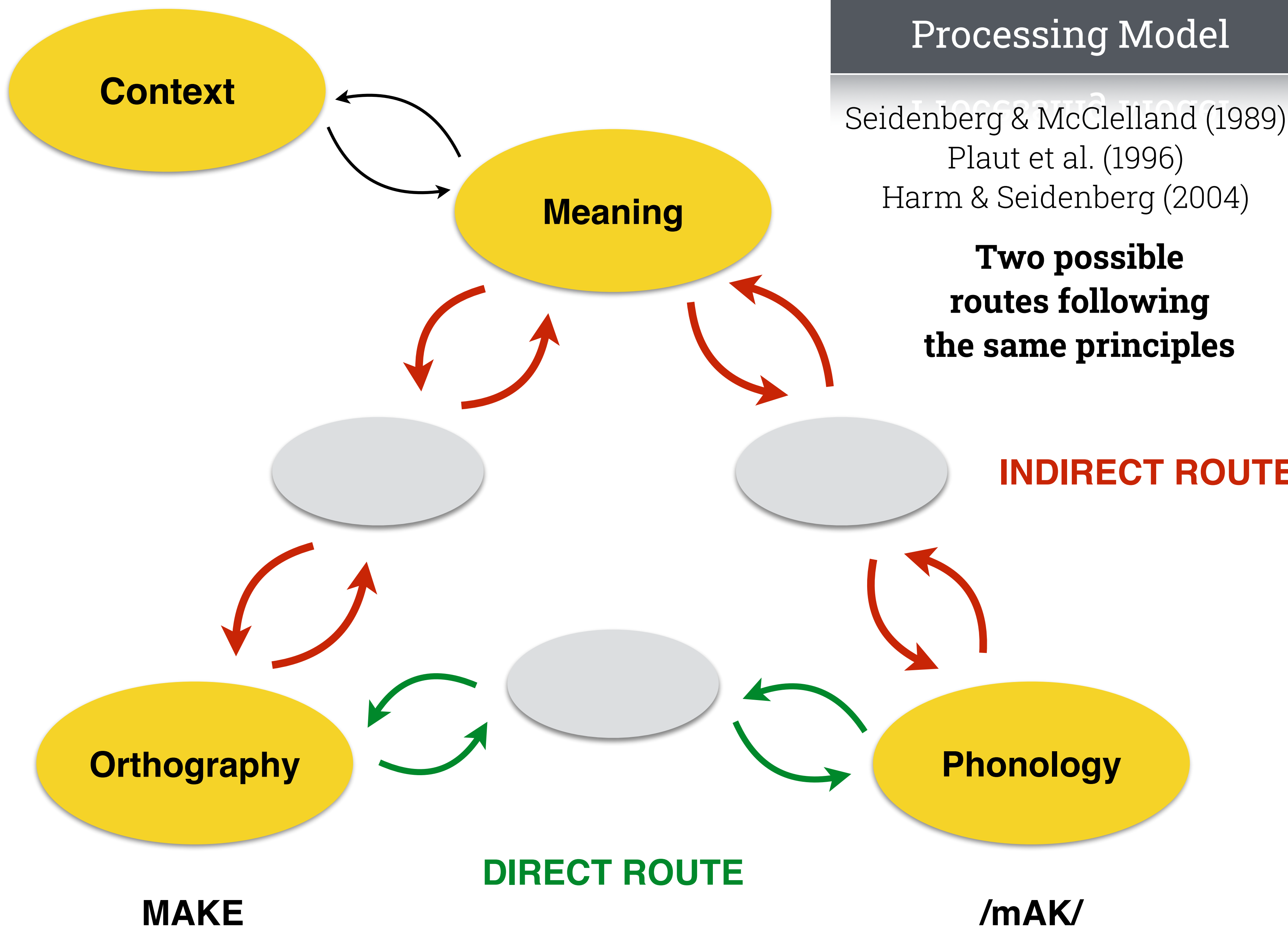


INDIRECT ROUTE

Parallel Distributed Processing Model

Seidenberg & McClelland (1989)
Plaut et al. (1996)
Harm & Seidenberg (2004)

Two possible routes following the same principles



THREE MODELS

Parallel Distributed Processing (PDP)

Dual-Route Cascaded (DRC)

Connectionnist Dual Process + (CDP+)

Dual-Route

- **Not new ! Marshall & Newcombe (1973)**

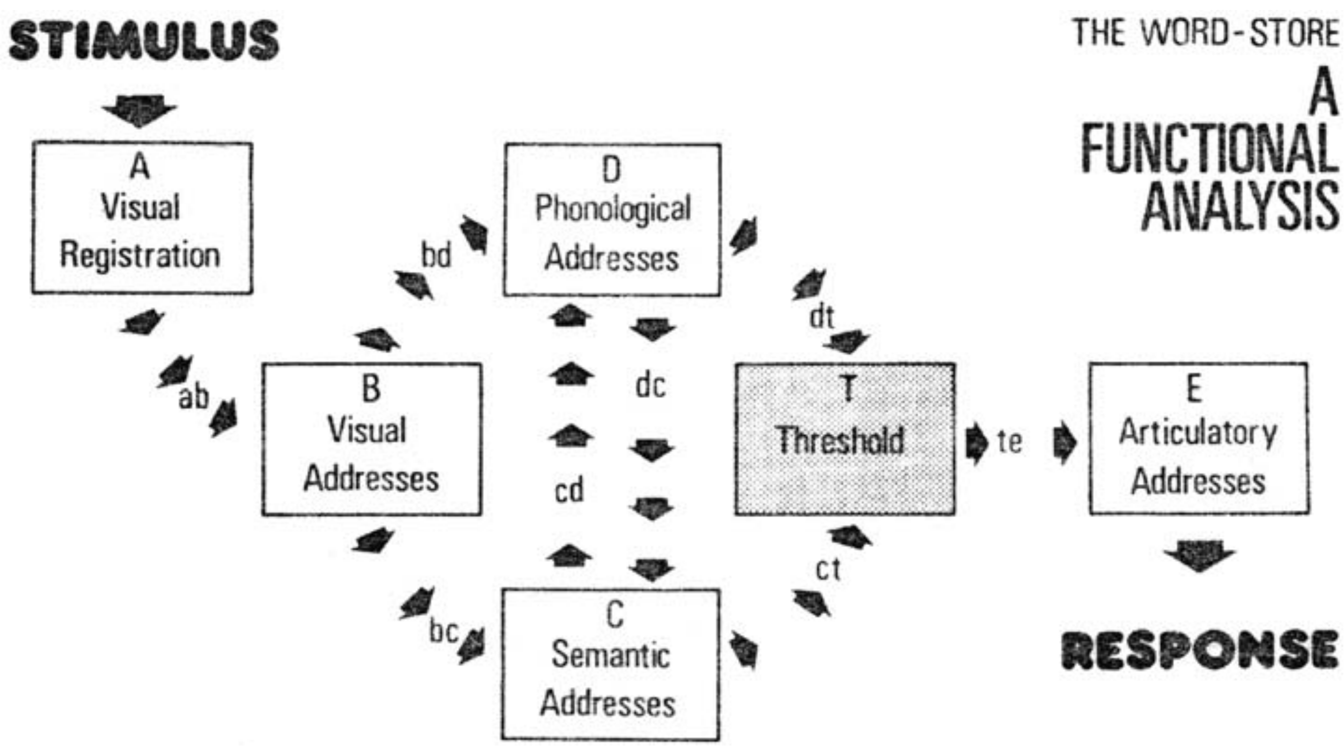
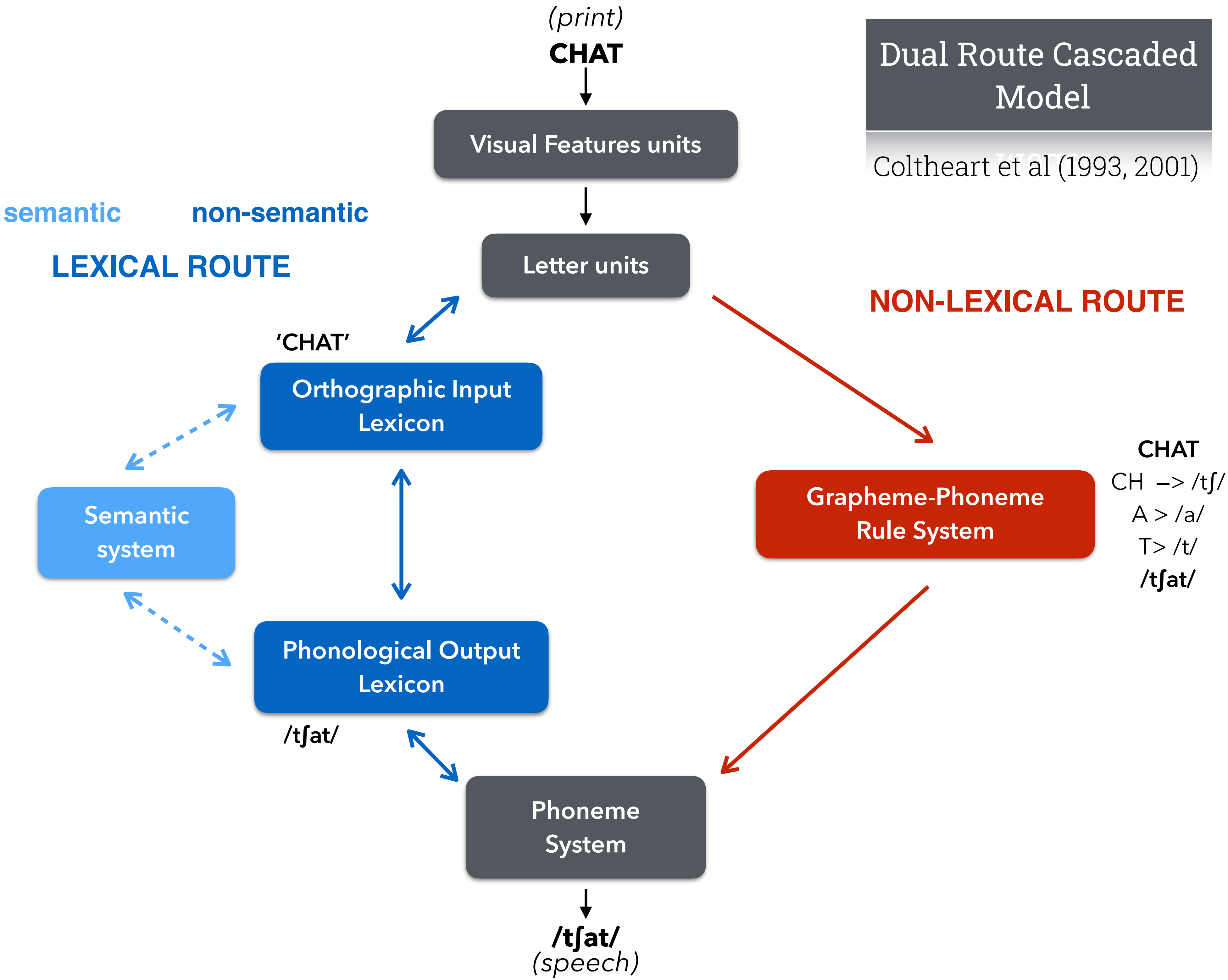


Figure 1. *A dual-route model of reading aloud (Marshall & Newcombe, 1973).*

Dual Route Cascaded Model

Coltheart et al (1993, 2001)



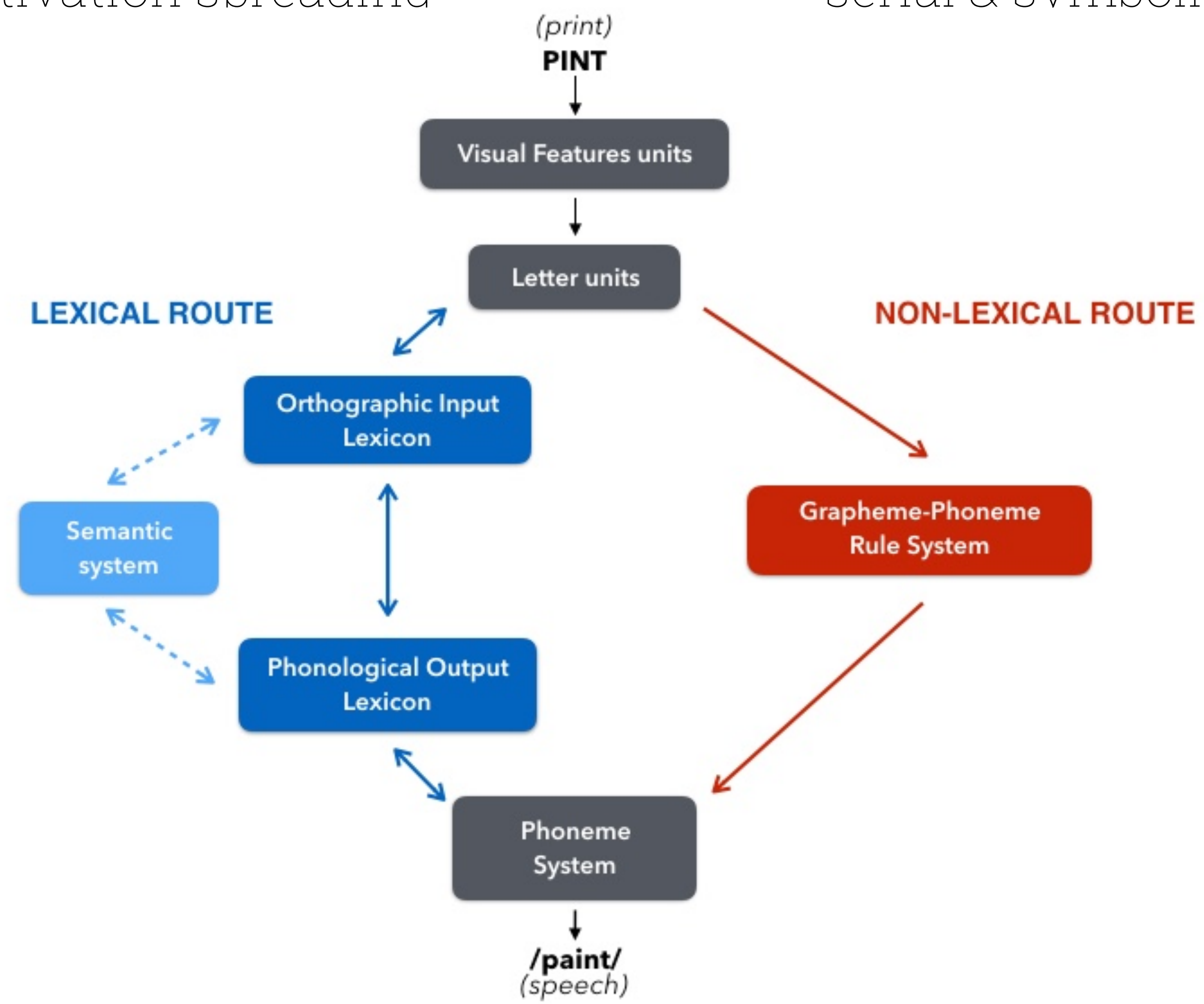
Two routes different and independent processing mechanisms

Lexical route

parallel & activation spreading

Non-Lexical route

serial & symbolic



THREE MODELS

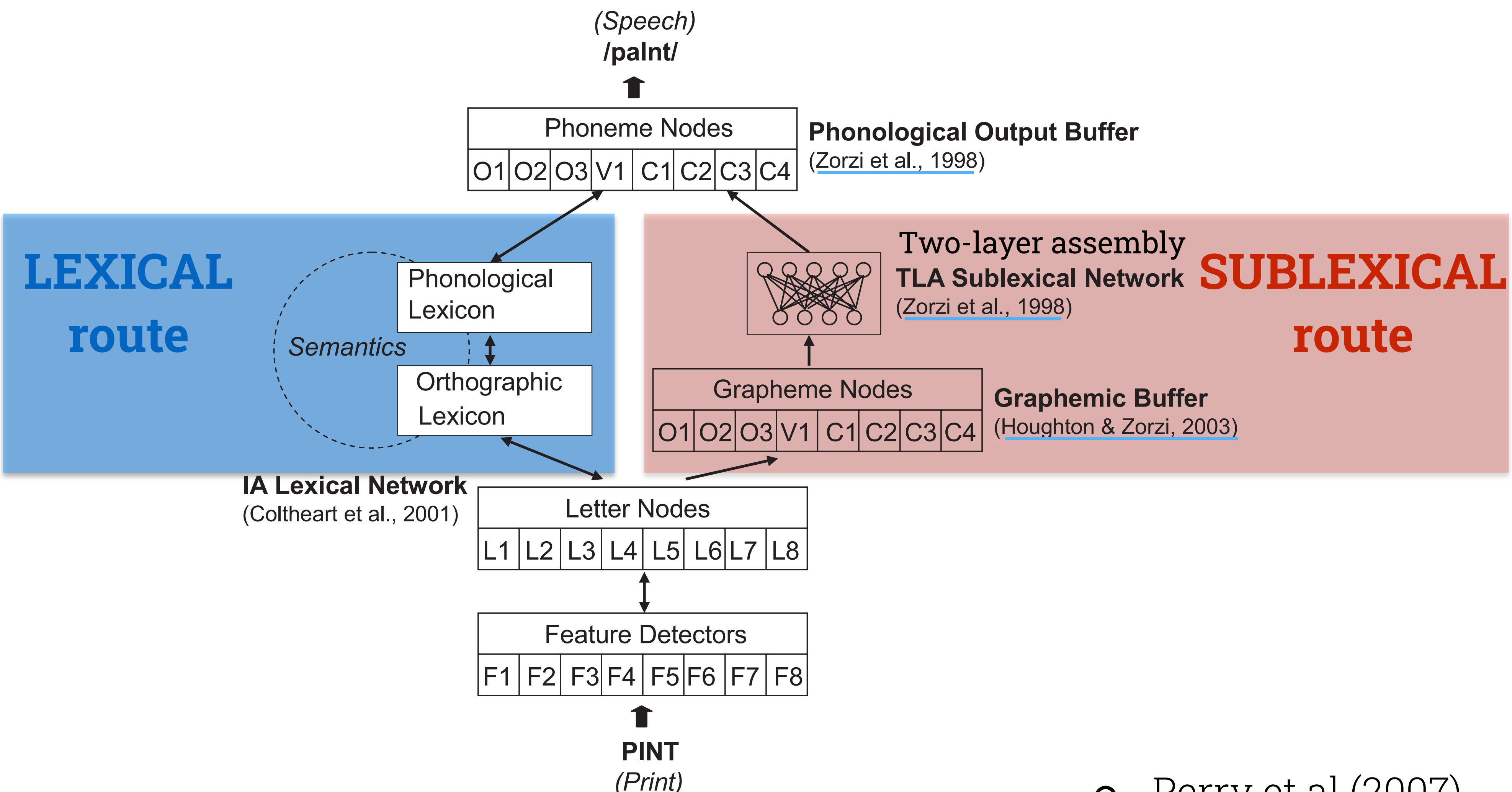
Parallel Distributed Processing (PDP)

Dual-Route Cascaded (DRC)

Connectionnist Dual Process + (CDP+)

Connectionist Dual Process + (CDP+)

Hybrid model : CDP model with DRC's lexical route and a connectionist approach of the non lexical route.



- Perry et al (2007)

How to evaluate them ?

- **Accuracy and Speed of reading**

  for models :
  - phonological output
  - number of processor cycles to produce it

- **Evaluate their simulation performance of some effects observed in human reading**

- **Evaluate theoretical choices that lead to their design and their objectives**

PERFORMANCES AT SIMULATING SOME EFFECTS

2

Simulating effects observed in human reading

New List of Benchmark Effects

Name of effect	Benchmark data set	Description	Triangle	DRC	CDP+
Frequency	Jared (2002, Experiment 2) Weekes (1997)	High-frequency words are faster/more accurate than low-frequency words.	+	+	+
Lexicality	McCann and Besner (1987) Weekes (1997)	Words are faster/more accurate than pseudowords.	+	+	+
Frequency × Regularity	Paap and Noel (1991) Jared (2002, Experiment 2)	Irregular words are slower/less accurate than regular words. Jared (2002) reported no interaction with frequency.	−	+	+
Word consistency	Jared (2002, Experiment 1)	Inconsistent words are slower/less accurate than consistent words. The size of the effect depends on the friend–enemy ratio.	+	−	+
Nonword consistency	Andrews and Scarratt (1998)	Nonword pronunciations show graded consistency effects; that is, people do not always use the most common grapheme–phoneme correspondences.	−	−	+
Length × Lexicality	Weekes (1997) Ziegler et al. (2001)	Nonword naming latencies increase linearly with each additional letter.	−	+	+
Position of irregularity	Rastle and Coltheart (1999)	The size of the regularity effect is bigger for words with first position irregularities (e.g., *chef*) than for words with second- or third-position irregularities.	−	+	+
Body neighborhood	Ziegler et al. (2001)	Words with many body neighbors are faster/more accurate than words with few body neighbors.	−	−	+
Masked priming	Forster and Davis (1991)	Words preceded by an onset prime are faster/more accurate than words preceded by unrelated primes.	?	+	+
Pseudohomophone advantage	McCann and Besner (1987) Reynolds and Besner (2005)	Nonwords that sound like real words (e.g., *bloo*) are faster/more accurate than orthographic controls.	+	+	+
Surface dyslexia	Patterson and Behrmann (1997)	Patient MP showed specific impairment of irregular word reading, which was modulated by the consistency ratio of the words.	+	−	+
Phonological dyslexia	Derouesné and Beauvois (1985)[a]	Patient LB showed specific impairment of nonword reading, which was reduced when nonwords were orthographically similar pseudohomophones.	+[b]	+	+
Large-scale databases	Spieler and Balota (1997) Balota and Spieler (1998)	Naming latencies of the model were regressed onto the average naming latency of each item in large-scale databases containing thousands of items.	−	−	+

Simulating effects observed in human reading

Frequency effect

Nonwords reading

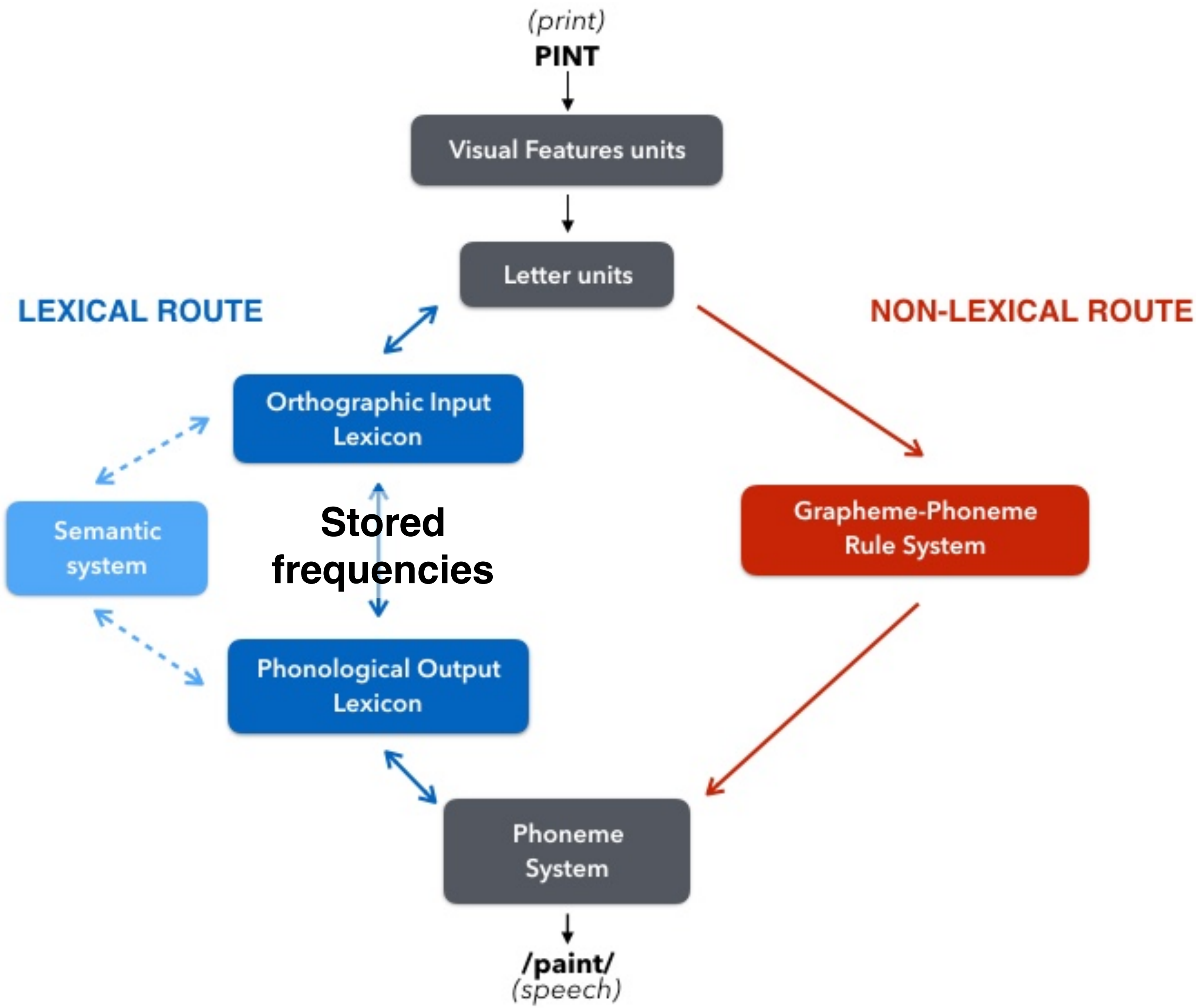
Regularity and consistency

Phonological / Surface dyslexia

Effect of frequency

More frequent words are read faster/ more accurately than less frequent ones

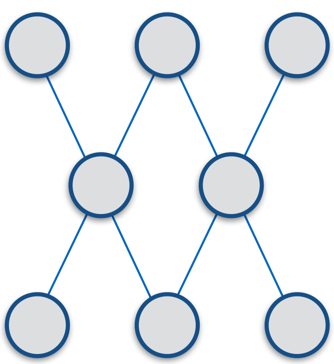
- Considered as an evidence of the existence of a dictionary
- DRC : lexical access speed depends on frequency, these frequencies are stored in the lexicon


(print)
PINT
Visual Features units
Letter units
LEXICAL ROUTE
NON-LEXICAL ROUTE
Orthographic Input Lexicon
Semantic system
Stored frequencies
Grapheme-Phoneme Rule System
Phonological Output Lexicon
Phoneme System
/paint/
(speech)

Effect of frequency

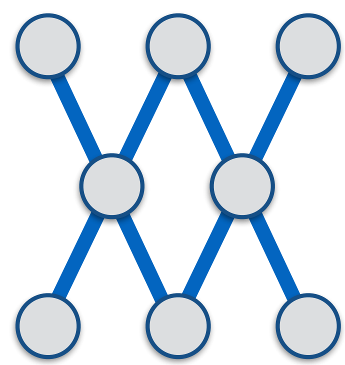
More frequent words are read faster/ more accurately than less frequent ones

- Considered as an evidence of the existence of a dictionary
- DRC : lexical access speed depends on frequency, these frequencies are stored in the lexicon
- PDP : frequency effect emerges from independent principles, it is a consequence of the learning procedure



non frequent words

more frequent words
have stronger connections
and are then processed faster



frequent words

Simulating effects observed in human reading

New List of Benchmark Effects

Name of effect	Benchmark data set	Description	Triangle	DRC	CDP+
Frequency	Jared (2002, Experiment 2) Weekes (1997)	High-frequency words are faster/more accurate than low-frequency words.	+	+	+
Lexicality	McCann and Besner (1987) Weekes (1997)	Words are faster/more accurate than pseudowords.	+	+	+
Frequency × Regularity	Paap and Noel (1991) Jared (2002, Experiment 2)	Irregular words are slower/less accurate than regular words. Jared (2002) reported no interaction with frequency.	−	+	+
Word consistency	Jared (2002, Experiment 1)	Inconsistent words are slower/less accurate than consistent words. The size of the effect depends on the friend–enemy ratio.	+	−	+
Nonword consistency	Andrews and Scarratt (1998)	Nonword pronunciations show graded consistency effects; that is, people do not always use the most common grapheme–phoneme correspondences.	−	−	+
Length × Lexicality	Weekes (1997) Ziegler et al. (2001)	Nonword naming latencies increase linearly with each additional letter.	−	+	+
Position of irregularity	Rastle and Coltheart (1999)	The size of the regularity effect is bigger for words with first position irregularities (e.g., *chef*) than for words with second- or third-position irregularities.	−	+	+
Body neighborhood	Ziegler et al. (2001)	Words with many body neighbors are faster/more accurate than words with few body neighbors.	−	−	+
Masked priming	Forster and Davis (1991)	Words preceded by an onset prime are faster/more accurate than words preceded by unrelated primes.	?	+	+
Pseudohomophone advantage	McCann and Besner (1987) Reynolds and Besner (2005)	Nonwords that sound like real words (e.g., *bloo*) are faster/more accurate than orthographic controls.	+	+	+
Surface dyslexia	Patterson and Behrmann (1997)	Patient MP showed specific impairment of irregular word reading, which was modulated by the consistency ratio of the words.	+	−	+
Phonological dyslexia	Derouesné and Beauvois (1985)[a]	Patient LB showed specific impairment of nonword reading, which was reduced when nonwords were orthographically similar pseudohomophones.	+[b]	+	+
Large-scale databases	Spieler and Balota (1997) Balota and Spieler (1998)	Naming latencies of the model were regressed onto the average naming latency of each item in large-scale databases containing thousands of items.	−	−	+

Reading nonword

A nonword is a a sequence of letter that is not a word but that is somehow readable = that respects the orthographic/phonotactic rules of the language

PLONE

PARTOR

CHITAP

~~WHJKOIT~~

~~PGIDBETX~~

Reading nonwords

Described as a big problem for PDP models (SM89)

How do you evaluate simulated pronunciations as there are no attested pronunciations for nonwords ?

- DRC : respecting GPC rules

- PDP : phonemic sequence contains subcomponents which pronunciation can be found in real words

And experimental data ?

Reading nonwords

- **Pritchard (2012)**

  Reading 412 nonword by 47 australian participants

  DRC > 73,5% of correspondance

  CDP+ adjusted > 20% of correspondance

  But all DRC simulated pronunciations are regular !

Table 2
Classification of Pronunciations Given by Experiment Participants and by Each Model

Classification	Experiment participants	DRC	CDP+	CDP.50
Regular	53.0	100.0	0.0	9.0
Vowel difference	20.3	0.0	36.7	35.7
Dropped phoneme	1.2	0.0	11.4	5.1
Extra phoneme	4.0	0.0	7.3	5.1
s/z coda difference	4.0	0.0	10.9	11.9
Consonant difference	5.1	0.0	16.7	20.1
Other	12.4	0.0	17.0	13.1

Reading nonwords

- **Perry et al (2014)**

Reading 70 (28) nonword by 32 french participants

Final consonants are most of the time not pronounce in French

> GPC rule is then not to pronounce them

But participants were often pronouncing them in nonword

> correctly reproduced by CDP++ , not by DRC

DRC is quantitatively better but qualitatively weaker

(problems with nonwords that are not pronounced regularly)

Spelling-sound regularity

A word has a regular prononciation if its pronunciation can be **derived entirely from the grapho-phonemic conversion rules,** which are based on the statistically most plausible pronunciation.

IRREGULAR	**REGULAR**
VILLE	FILLE, QUILLE, BILLE
PINT	MINT, TINT ..
HAVE	GAVE, SAVE, RAVE
CHEF	CHOP, CHALLENGE

Spelling-sound regularity effect

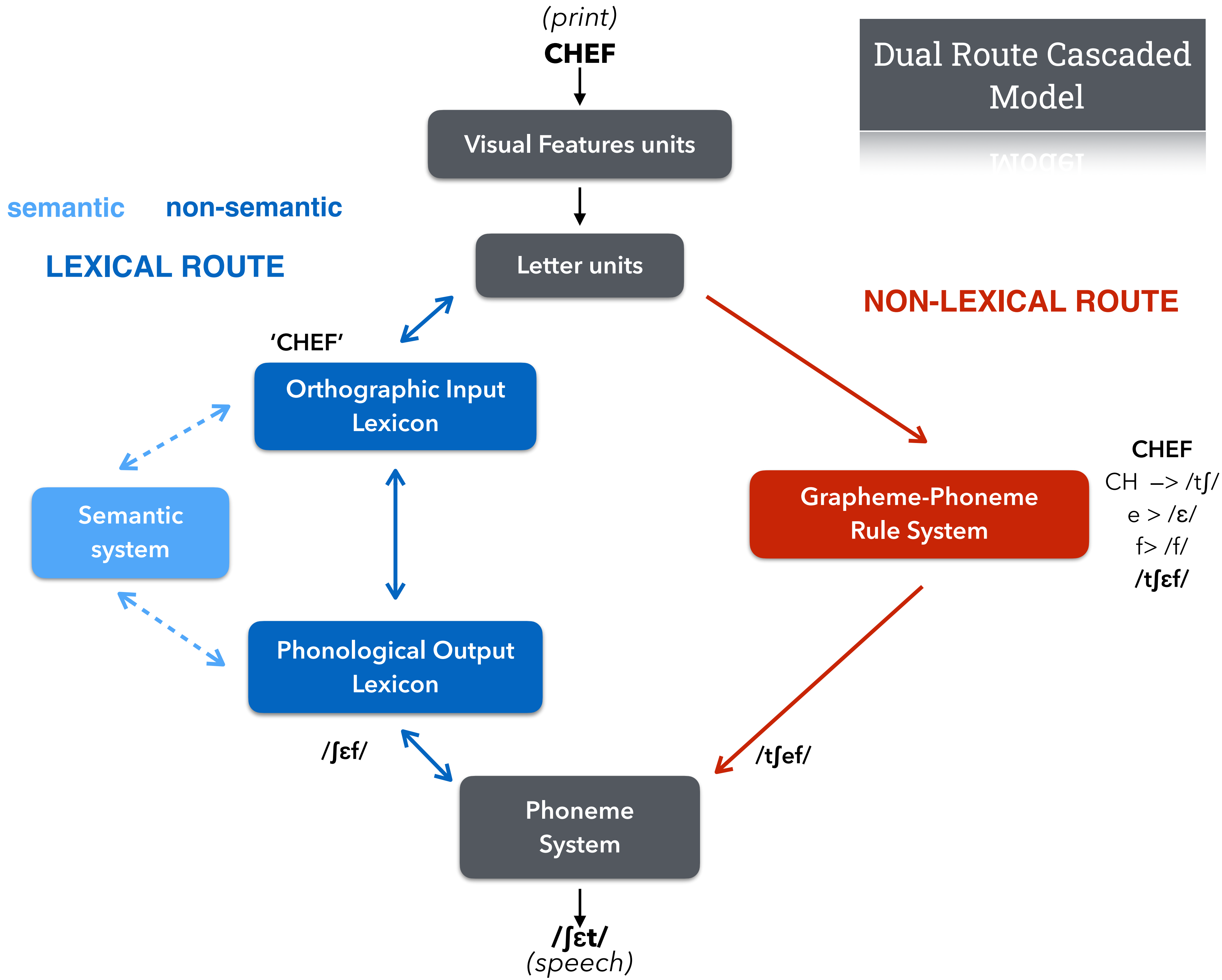
Words with a non regular pronunciation are read more slowly than those with a regular one

PDP : A side effect of frequency effect

DRC : A conflict of output between the two routes

CDP+ : A combination of both

Dual Route Cascaded Model



Rime consistent pronunciation

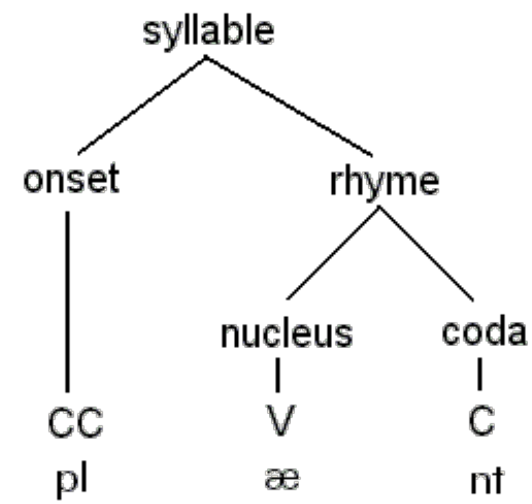
A word is consistent if its rime have the same pronunciation in all other words with the same rime

- **Consitent**

-EAN : BEAN, LEAN, JEAN, MEAN, DEAN

-OND : BOND, FROND, POND, FOND

-ONT : FRONT, PONT, VONT



- **Non-consistent**

PINT	MINT, TINT ..
HAVE	GAVE, SAVE, RAVE
VILLE	FILLE, QUILLE, BILLE

Regularity - Consistency interaction

Questions about influence of each effect

(Glushko, 1979, Jared, McRae & Seidenberg, 1990)

Two independent effects

(Cortese & Simpson, 2000 ; Jared,,1997)

PDP predicts an important effect of consistency

DRC predicts an important effect of regularity

> Observation : Important effect of consistency

(Cortese & Simpson, 2000 ; Jared,2002)

DRC fails to simulate consistency effects for nonword

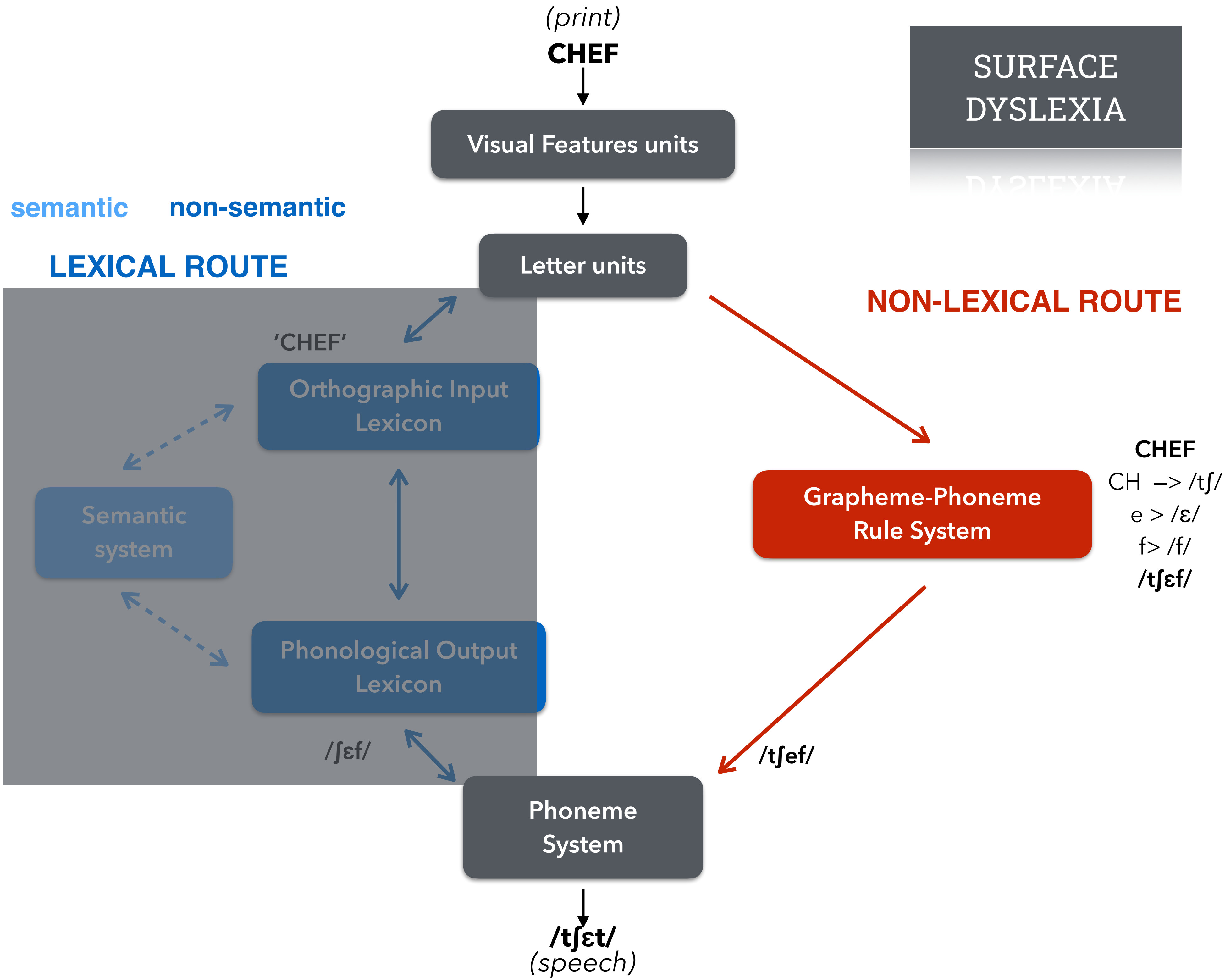
(Zevin & Seidenberg, 2006)

Dyslexia

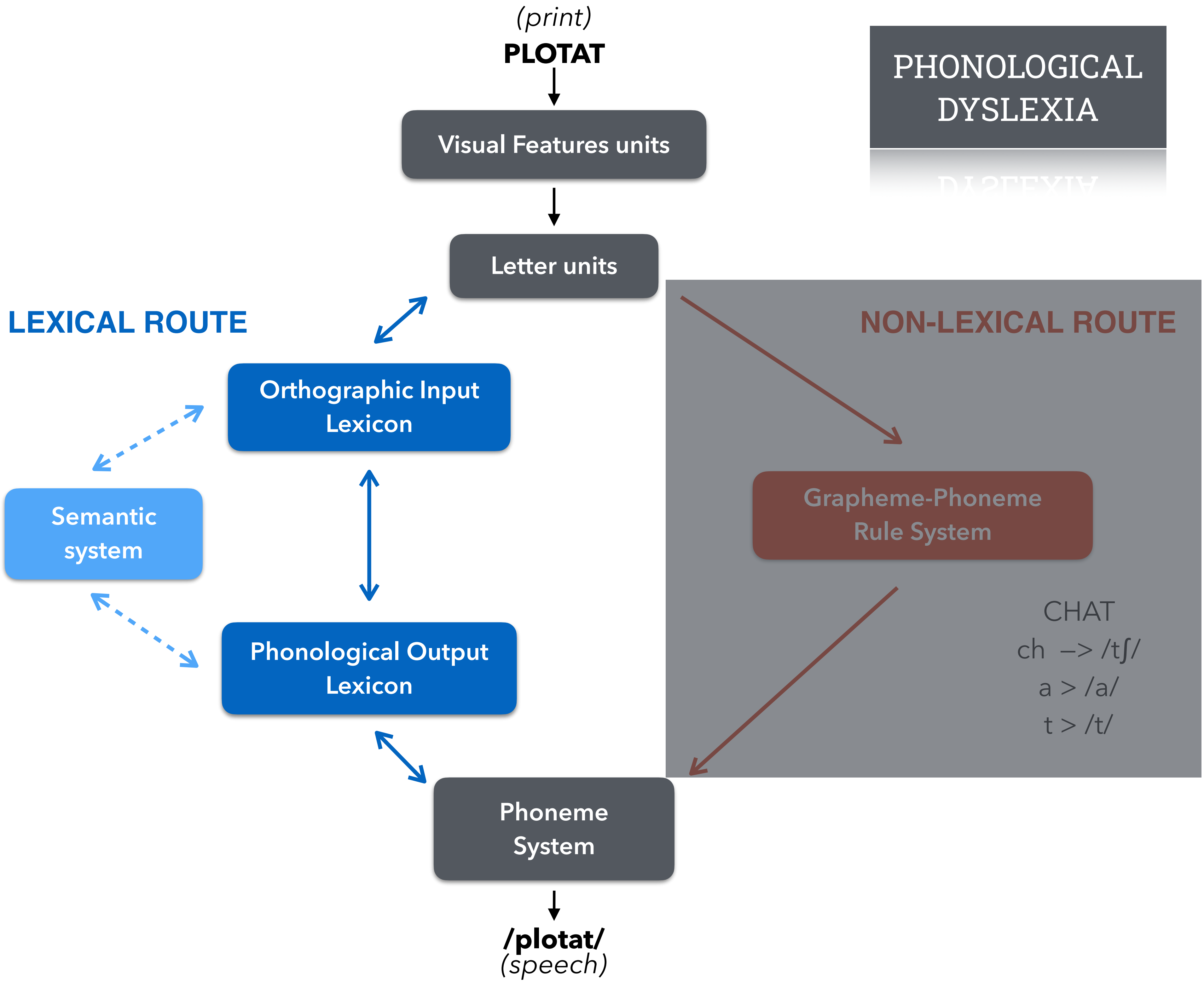
Why having two routes ?

To represent a double dissociation between surface and phonological dyslexia

	SURFACE DYSLEXIA	PHONOLOGICAL DYSLEXIA
Regular words	✓	✓
Non-words New words	✓	✗
Irregular words	✗	✓


SURFACE DYSLEXIA
(print)
CHEF
Visual Features units
Letter units
semantic
non-semantic
LEXICAL ROUTE
'CHEF'
Orthographic Input Lexicon
Semantic system
Phonological Output Lexicon
/ʃɛf/
NON-LEXICAL ROUTE
Grapheme-Phoneme Rule System
CHEF
CH –> /tʃ/
e > /ɛ/
f> /f/
/tʃɛf/
/tʃef/
Phoneme System
/tʃɛt/
(speech)

PHONOLOGICAL DYSLEXIA



THEORETICAL APPROACHES

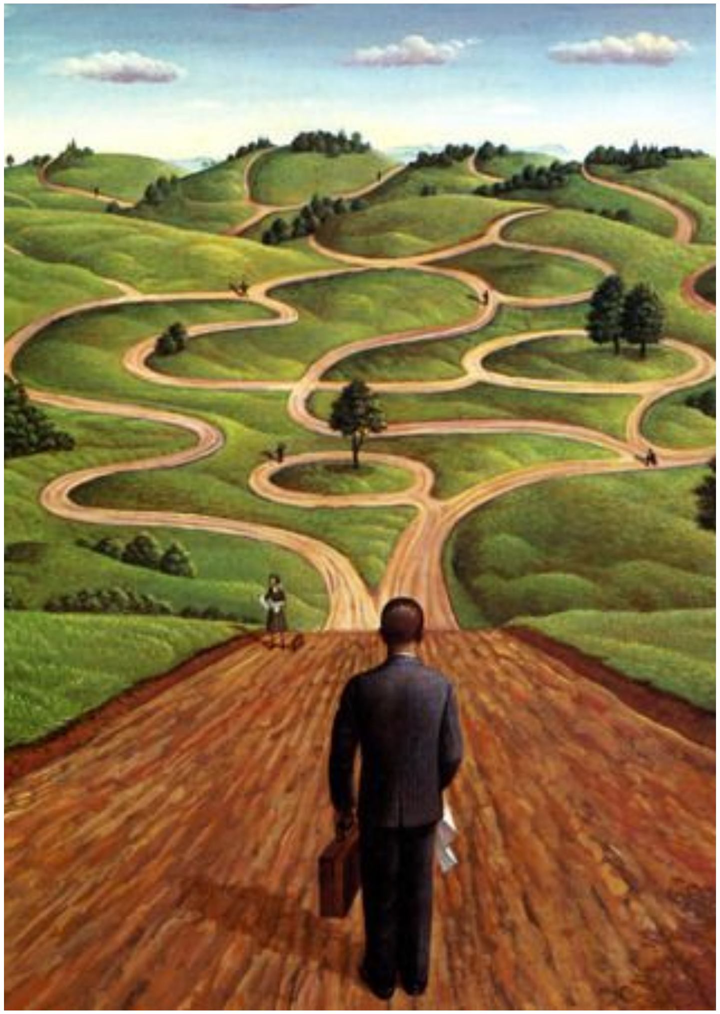
3

Main differences

- **number of routes and processes**
- **learning procedure or not**
- **theory building versus Data fitting**
- parallel or serial processing
- distributed or local representations

Number of routes & double dissociation

- Double dissociation > 2 routes
- But one route/processing damaged in two different manners could display the same dissociation (~PDP simulations)
- An observed difference/dissociation is not always directly transposed in the processing architecture. (Davies, 2010)



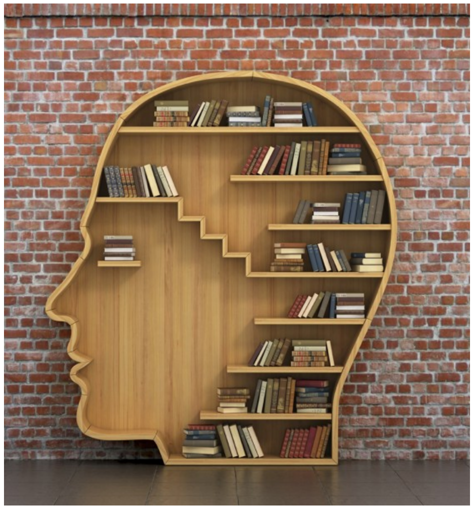
Learning procedure or not

- PDP : a general learning algorithm
  (retropropagation of error)
- CDP+ : a learning algorithm for the non-lexical route
  (delta learning rule)
- DRC : no learning, fixed conversion rules, models critical values are entered manually in the model by researchers based on external statistics

Having a learning algorithm potentially allows :

- to simulate reading development
- to simulate frequency trajectory effects
- to have a independent explanation for some effect

Opposite visions of modelling



PDP > **Theory building**

using plausible processing given what we know about neural networks and learning

models are tools

DRC and CDP+ > **Data fitting**

models are the aim of the research, fitting the data is more important than a detailed explanation of underlying mechanisms

Danger of overfitting the data

Which new hypotheses can emerge ?

Explanatory adequacy vs. Descriptive adequacy

LIMITS AND CONCLUSION

4

Methodological limitations

Factorial approach has some caveats :

- Different experimental results
- Not a unique corpus of word to evaluate different effects
- Changing model's parameters for each effect
- Dichotomy of continuous variable (worse for non-linear variable)

Lists problems

- Impossible to create list of words that vary only for one parameter
- Effect of list composition
- Interaction between items in the list

Participants profiles

- Age, reading experience, education, vocabulary size often not taken into account

Some new possibilities

Mega studies : English Lexicon Project (Balota et al, 2007)

using big database rather than different little data collections

Mixed models of linear regression rather than ANOVA

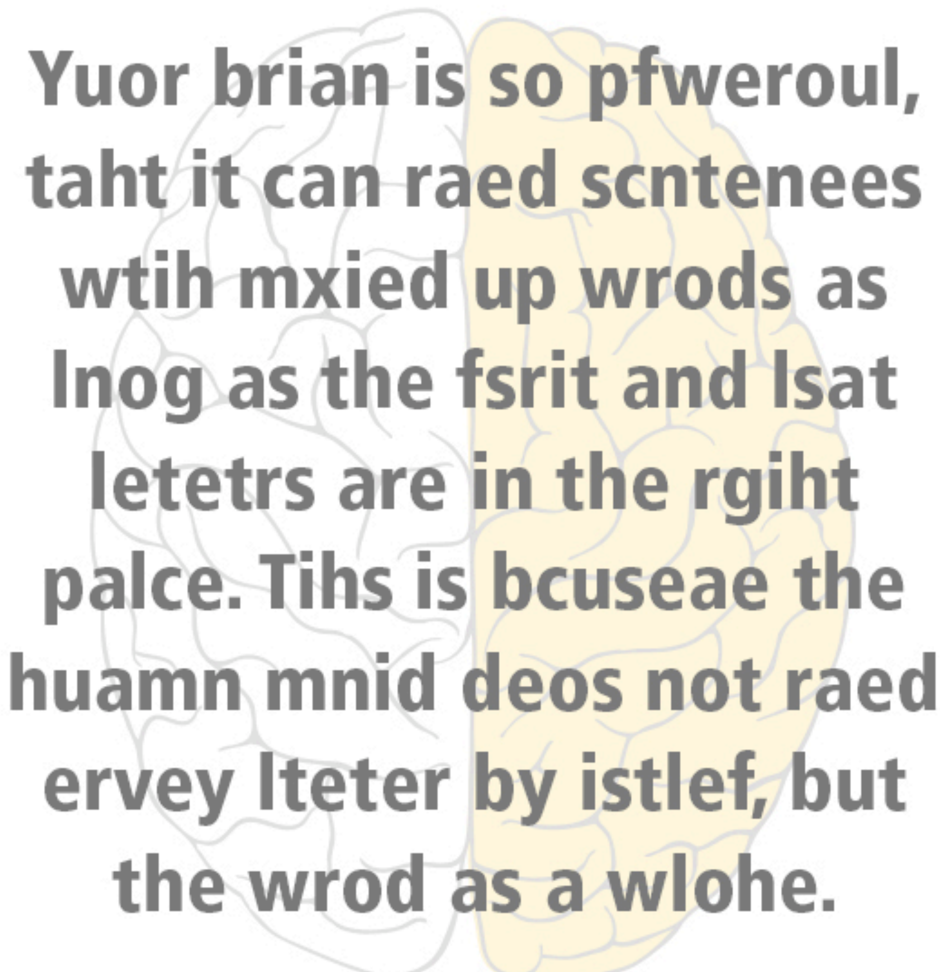
isolate words and participants variance

cannot change parameters for each effect !

Analysis of reading time distributions rather than just central tendencies

Some limits

- fix encoding of letters positions



Some limits

- fix encoding of letters positions
- monosyllabic words
- isolated words

La Liseuse (1872) Monet

Some limits

- fix encoding of letters positions
- monosyllabic words
- isolated words
- alphabetic languages only

Some limits

- fix encoding of letters positions
- monosyllabic words
- isolated words
- alphabetic languages only
- differences between languages ?

Different orthographic transparency

+ ↑ Finnish / Spanish

French / English

— Icelandic

Some limits

- fix encoding of letters positions
- monosyllabic words
- isolated words
- alphabetic languages only
- differences between languages ?
- inter-individual variation

Some limits

- fix encoding of letters positions
- monosyllabic words
- isolated words
- alphabetic languages only
- differences between languages ?
- universal model of reading ?

BEHAVIORAL AND BRAIN SCIENCES (2012) **35**, 263–329
doi:10.1017/S0140525X11001841

Towards a universal model of reading

Ram Frost
Department of Psychology, The Hebrew University, Jerusalem 91905, Israel, and Haskins Laboratories, New Haven, CT 06511
frost@mscc.huji.ac.il
http://psychology.huji.ac.il/en/?cmd=Faculty.113&letter=f&act=read&id=42~frost/

Reading aloud is not just recognizing words

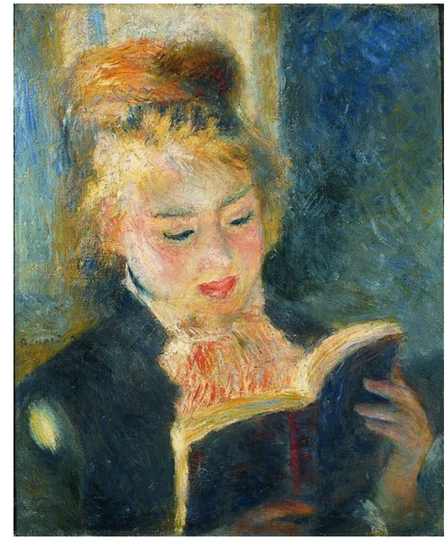
- Those experience in reading aloud don't just measure Visual Word Recognition, but also processes regarding phonological planning and motor programmation

- Neuroimaging and specifically ERP could be a good method to study recognition per se.

Conclusion

- **different simulations performances**
  - frequency effect
  - nonword reading
  - consistency and regularity
  - surface and phonological dyslexia
- **two main theoretical approaches**
  - distributed or local representations
  - number of routes : 1 or 2
  - parallel versus serial processing
  - applying rules or learning algorithm building connections patterns

Conclusion


La Liseuse (1890) Renoir

- **Parallel Distributed Processing Models**
  - processing principles that are close to supposed cerebral functionning
  - fewer effects represented, but often with a richer explanation
- **Dual-Route Cascaded Model**
  - more effects represented (except consistency effect)
  - easy to understand, but what new hypotheses ?
  - no learning procedure, fixed rules
- **CDP+**
  - best to simulate different effects
  - still concerns about mixing different principles

Conclusion

- No model has a clear superiority given their different approaches and objectives
- Methodological improvement for experimental data acquisition and models evaluation
- A better understanding of the reading cerebral architecture and deeper knowledge of orthographic processing could help future direction of research and maybe support one architecture
- All of this behind a single word !

The Chaos of English pronunciation

https://www.youtube.com/watch?v=1edPxKqiptw

Thanks

My committee :
Laura Gonnerman, Shari Baum and Karsten Steinhauer

Sara Colby, AJ Orena and Deirdre Truesdale
for their constant support

Questions

La Liseuse, Picasso (1937)

Pigeons can do it ! (?)

Stop everything! It turns out pigeons can read

Don't underestimate the pigeon -- they might look a bit dim but it turns out these birds can actually read.

Lead story

New Research shows Pigeons can be taught to Read, just like Humans

The pigeon can learn to distinguish between real words and non-words by looking at the letter combinations

AMERICA

Can Pigeons Spell? New Study Suggests They Can Recognize Words

September 22, 2016 · 4:00 PM ET

A novel theory suggests that orthographic processing is the product of neuronal recycling, with visual circuits that evolved to code visual objects now co-opted to code words. Here, we provide a litmus test of this theory by assessing whether pigeons, an organism with a visual system organizationally distinct from that of primates, code words orthographically. Pigeons not only correctly identified novel words but also display the hallmarks of orthographic processing, in that they are sensitive to the bigram frequencies of words, the orthographic similarity between words and nonwords, and the transposition of letters. These findings demonstrate that visual systems neither genetically nor organizationally similar to humans can be recycled to represent the orthographic code that defines words.

SCIENCE

If You Can Read This, You May Be A Pigeon

Some birds, it turns out, can be taught to read. Or at least spell-check.

09/20/2016 01:41 pm ET

Pigeons can do it ! (?)

Orthographic processing in pigeons (*Columba livia*)

Damian Scarf[a,1], Karoline Boy[b], Anelisie Uber Reinert[b], Jack Devine[c], Onur Güntürkün[b], and Michael Colombo[a,1]

[a]Department of Psychology, University of Otago, Dunedin 9054, New Zealand; [b]Department of Psychology, Institute of Cognitive Neuroscience, Biopsychology, Ruhr-University Bochum, D-44780 Bochum, Germany; and [c]Department of Physics, University of Otago, Dunedin 9054, New Zealand

Edited by Dale Purves, Duke University, Durham, NC, and approved July 29, 2016 (received for review May 17, 2016)

Learning to read involves the acquisition of letter–sound relationships (i.e., decoding skills) and the ability to visually recognize words (i.e., orthographic knowledge). Although decoding skills are clearly human-unique, given they are seated in language, recent research and theory suggest that orthographic processing

An open question is whether animals with brain architectures and visual systems dissimilar to primates also display this sensitivity to the statistical properties of words. Indeed, the recycling hypothesis has been built with our brain in mind and, more specifically, the hierarchical organization of the ventral visual system.