

Computational approaches to mind and brain: How language is profoundly shaped by its neural substrate

Paul Smolensky

Cognitive Science Department
Johns Hopkins University

Faculty of Human Sciences

9 Dec 2014

Macquarie University

Cognitive Science comes of age

Cognitive Science: Characterize cognition – the function of the brain – precisely

A very exciting time for
Cogsci: transitioning
to a hard science!

- ◆ critical to technology
that changes the world
- ◆ critical to an established hard science

Artificial Intelligence – AI

Watson

2010

Cognitive Science comes of age

Cognitive Science: Characterize cognition – the function of the brain – precisely

A very exciting
Cogsci: time
to a hard

♦ critical to the
that change

Artificial Int

Watson

Watson: IBM's AI system beat 2 all-time human champions in the classic American quiz-game show *Jeopardy!* (2011)



Cognitive Science comes of age

Cognitive Science: Characterize cognition – the function of the brain – precisely

A very exciting
Cogsci: time
to a hard

♦ critical to t
that chang

Artificial Int

Watson

Watson: IBM's AI system beat 2 all-time human champions in the classic American quiz-game show *Jeopardy!* (2011)

Input: English text; output: English speech

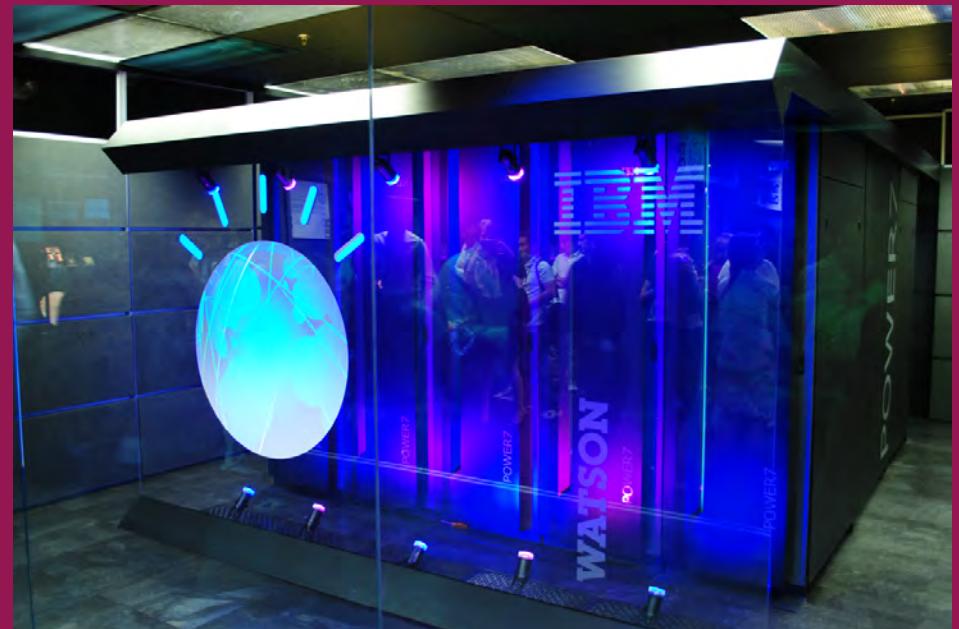
William Wilkinson's “An Account of the Principalities of Wallachia and Moldavia” inspired this author's most famous novel.

Who is Bram Stoker?

2,880 POWER7 processor cores

16 TB RAM [\$3M]

500 GB ~1M books/sec



Cognitive Science comes of age

Cognitive Science: Characterize cognition – the function of the brain – precisely

A very exciting time for
Cogsci: transitioning
to a hard science!

- ◆ critical to technology
that changes the world

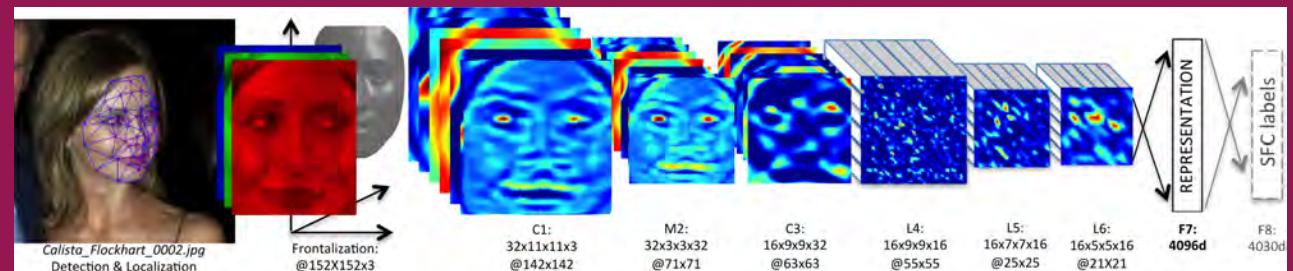
Artificial Intelligence –

Watson
Deep Neural Networks

Deep Neural Network (DNN) capabilities, 12/2014



3755 Chinese characters; nearly human-level accuracy
human error rate: 3.87% vs. 4.21% for DNN



13,323 web photos of 5,749 celebrities; photo pairs:
same or different? 97.35% [Facebook AI Research,
Taigman et al. 2014]

Cognitive Science comes of age

Cognitive Science: Characterize cognition – the function of the brain – precisely

A very exciting time for
Cogsci: transitioning
to a hard science!

- ◆ critical to technology
that changes the world

Artificial Intelligence – A

Watson
Deep Neural Networks

Deep Neural Network (DNN) capabilities, 12/2014

Google: uses DNNs in its:
Android speech recognizer
Google Goggles
Google+ photo search service
...

Microsoft: Speech recognition service

“Nvidia builds CUDA GPU programming library for
machine learning – so you don't have to ...
Craft a deep neural network on a graphics chipset”

Cognitive Science comes of age

Cognitive Science: Characterize cognition – the function of the brain – precisely

A very exciting time for
Cogsci: transitioning
to a hard science!

- ◆ critical to technology
that changes the world

Artificial Intelligence – AI

Watson

Deep Neural Networks

2010

2010

Gradient Symbolic
Computing

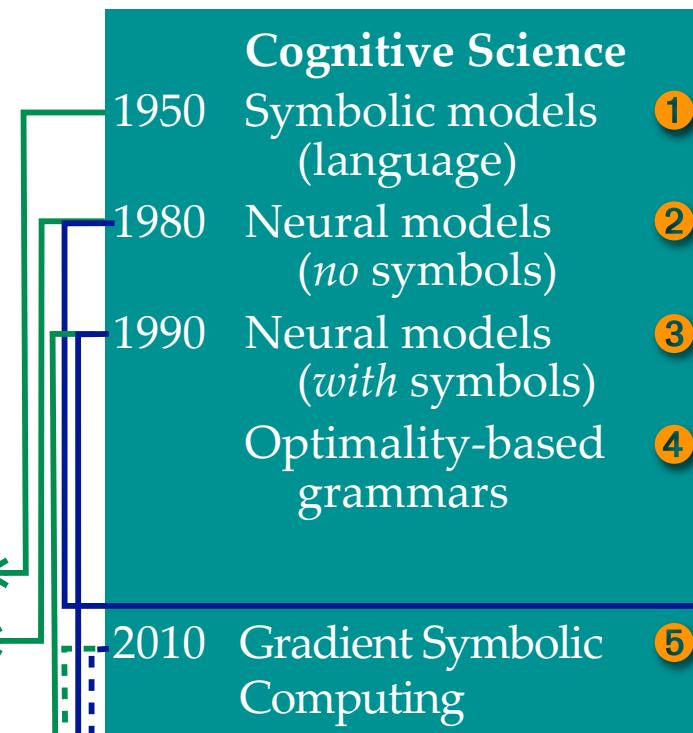
5

Brain reading

2000 gross functions

2010 single objects

2020 complex events



Today's topics

- ◆ critical to an established hard science

Cognitive Science comes of age

Cognitive Science: Computational approaches to mind and brain precisely

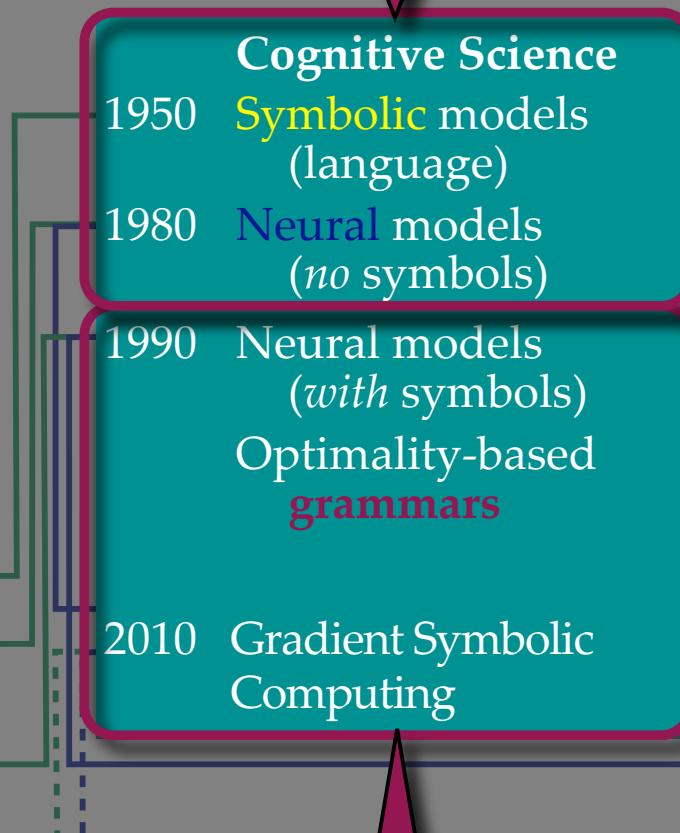
A very exciting time for
Cogsci: transitioning
to a hard science!

- ◆ critical to technology
that changes the world

Artificial Intelligence — AI

Watson
Deep Neural Networks

2010
2010



Today's topics

- ◆ critical to an established hard science

Brain reading

2000 gross functions
2010 single objects
2020 complex events

How language is profoundly shaped by its neural substrate

Cognitive Science comes of age

Cognitive Science: Characterize cognition – the function of the brain – precisely

A very exciting time for
Cogsci: transitioning
to a hard science!

- ◆ critical to technology
that changes the world

Artificial Intelligence – AI

Watson (IBM) 2010
Deep Neural Networks 2010

Cognitive Science		
1950	Symbolic models (language)	①
1980	Neural models (no symbols)	②
1990	Neural models (with symbols) Optimality-based grammars	③
2010	Gradient Symbolic Computing	④
		⑤

Today's topics

- ◆ critical to an established hard science

Brain reading

2000 gross functions
2010 single objects
2020 complex events

Symbolic cognitive models I: Reasoning

The notion at the heart of symbolic cognitive modeling has a long and distinguished pedigree: at least as far back as Aristotle. (4c. BCE)

- ◆ If I tell you, “it’s a law of science that *if p is true, then q is also true*”

and then I tell you, “*it so happens that p is true*”
then you can respond ... “*then q is true*”.

But you have no idea what p and q ARE. How can you know q is true?

It’s amazing that we can understand *if p is true, then q is also true* without knowing anything about p and q — understand it to reason with.

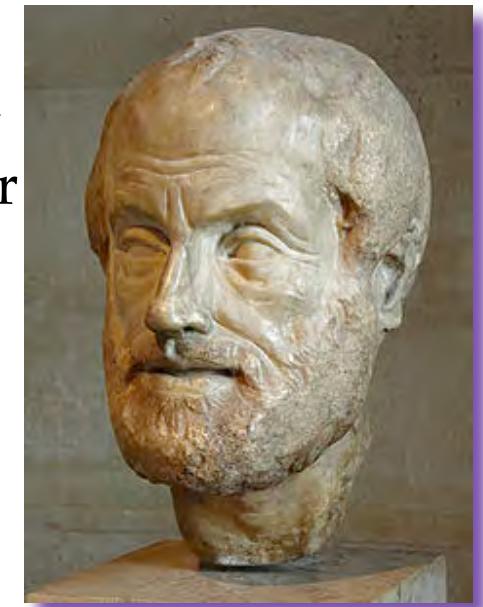
Humans are symbolic animals.

Our minds eat symbols for breakfast, lunch & dinner.

So do computers. Can we put *our mental symbolic world* into a computer?

That is the objective of *Symbolic Artificial Intelligence* — Symbolic AI

Its success is part of what brought us Watson.



Symbolic cognitive models 2: Language (grammar)

In language, like in reasoning, the historical foundation of symbolic cognitive modeling is deep. As far back as the Sanskrit grammarian Pāṇini (पाणिनि) in ancient India (4c BCE), symbolic rules have been used to state grammars with mathematical precision. (He had 3,959 rules just for constructing *words*.)

In the modern era (since 1950), a primary founder of the symbolic theory of grammar is Noam Chomsky.

In this approach, a grammar is a machine in a speaker's mind that manipulates symbols like *John*, *Noun*, *Verb Phrase*, ... to literally construct the sentences we produce as speakers.

The following is a concrete example from the part of grammar linguists call *phonology*: here, the rules govern how basic speech sounds like *a*, *m*, *k*, ... can be assembled into possible words.



These (few) gray slides make more technical points that are not critical for following the remainder of the talk.

Nerd slide — Symbolic grammar: phonology of Lardil

Lardil How to say ‘wooden axe’?

If an object:

muŋkumuŋkun = muŋkumuŋku-n
stem-suffix

If a subject:

muŋkumu

From what general knowledge
can the speaker compute this,
and all the other forms?



Nerd slide — Symbolic grammar: phonology of Lardil

Lardil How to say ‘wooden axe’?

If an object:

muŋkumuŋku = muŋkumuŋku-n
stem-suffix

If a subject:

muŋkumu

From what general knowledge can the speaker compute this, and all the other forms?

A grammar:

A sequence of rules that manipulate symbols

subject form:

muŋkumuŋku

muŋkumuŋku

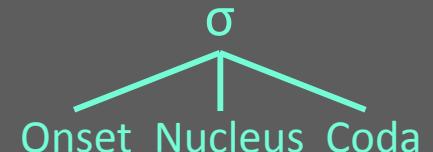
muŋkumuŋk

muŋkumuŋk

muŋkumuŋ

muŋkumuŋ

muŋkumu



Syllable-construction rules

Subject rule: delete final Vowel

Syllable-destructing rules:
No Nucleus ⇒ no σ

Delete unsyllabified segments

Syllable-destructing rules:
ŋ in Coda only if following k

Delete unsyllabified segments

à la Wilkinson 1983

Nerd slide — Symbolic grammar: phonology of Lardil

Lardil How to say ‘wooden axe’?

If an object:

muŋkumuŋku = muŋkumuŋku-n
stem-suffix

If a subject:

muŋkumu

From what general knowledge can the speaker compute this, and all the other forms?

A grammar:

A sequence of rules that manipulate symbols

subject form:

muŋkumuŋku

muŋkumuŋku

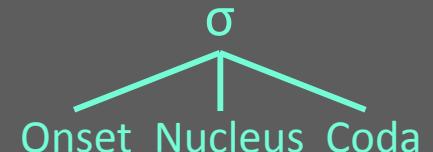
muŋkumuŋk

muŋkumuŋk

muŋkumuŋ

muŋkumuŋ

muŋkumu



Syllable-construction rules

Subject rule: delete final Vowel

Syllable-destructing rules:
No Nucleus ⇒ no σ

Delete unsyllabified segments

Syllable-destructing rules:
ŋ in Coda only if following k

Delete unsyllabified segments

à la Wilkinson 1983

Cognitive Science comes of age

Cognitive Science: Characterize cognition – the function of the brain – precisely

A very exciting time for
Cogsci: transitioning
to a hard science!

- ◆ critical to technology
that changes the world

Artificial Intelligence – AI

Watson (IBM) 2010
Deep Neural Networks 2010

Cognitive Science		
1950	Symbolic models (language)	①
1980	Neural models (no symbols)	②
1990	Neural models (with symbols) Optimality-based grammars	③
2010	Gradient Symbolic Computing	④
		⑤

Today's topics

- ◆ critical to an established hard science

Brain reading

2000 gross functions
2010 single objects
2020 complex events

Watson's inputs and outputs
are in English: the success of
symbolic linguistic theories was
also important in bringing us
Watson.

Cognitive Science comes of age

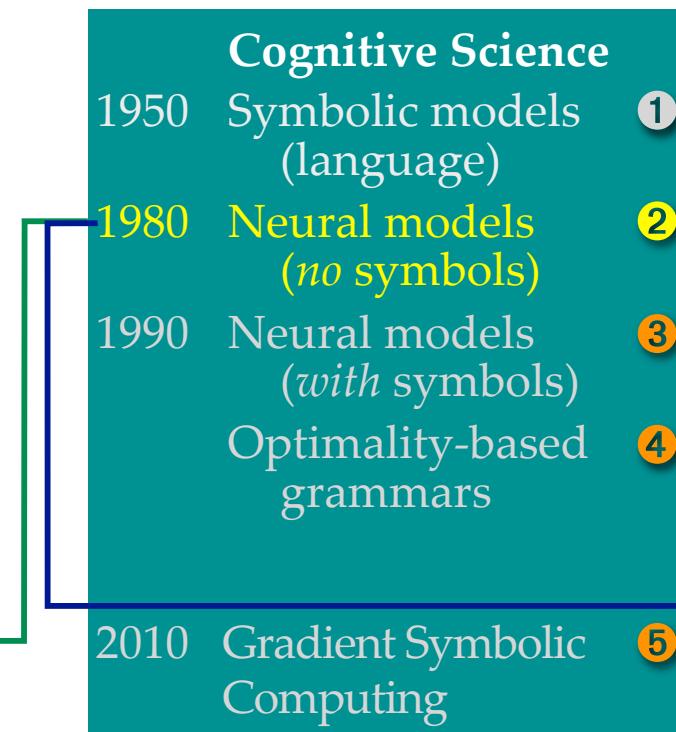
Cognitive Science: Characterize cognition – the function of the brain – precisely

A very exciting time for
Cogsci: transitioning
to a hard science!

- ◆ critical to technology
that changes the world

Artificial Intelligence – AI

Watson (IBM) 2010
Deep Neural Networks 2010



Today's topics

- ◆ critical to an established hard science

Brain reading

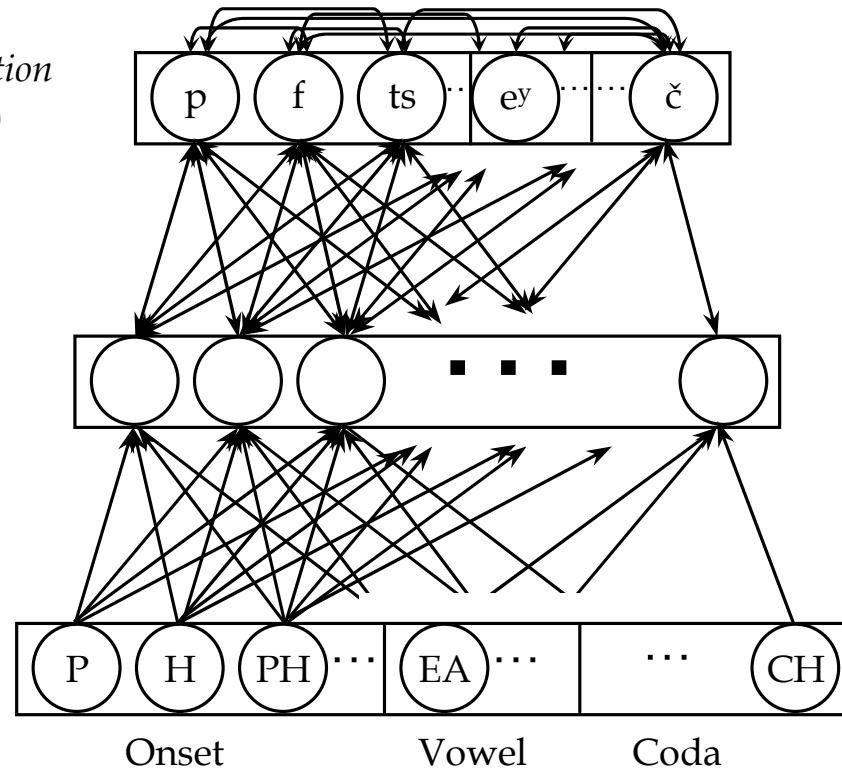
2000 gross functions
2010 single objects
2020 complex events

Neural models

Output: Pronunciation Segment string (61)

Hidden layer
(100)

Input: Orthography
Letter string (105)



Plaut, McClelland, Seidenberg &
Patterson (1996) *Psychological Review*

Used to explain detailed behaviors observed in the laboratory showing, e.g., that words are produced more slowly when there are a number of other frequent words with similar spelling but dissimilar pronunciations:

GOWN (vs. GROWN, BLOWN, ...)
slower than

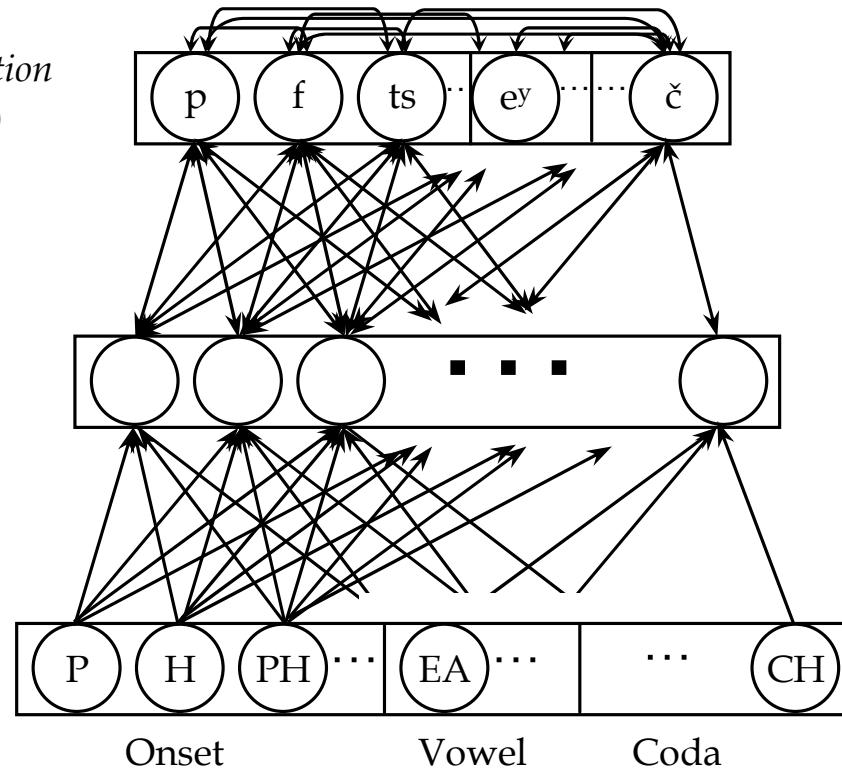
MINT (vs. PINT)

Neural models

Output: Pronunciation Segment string (61)

Hidden layer
(100)

Input: Orthography
Letter string (105)



Plaut, McClelland, Seidenberg &
Patterson (1996) *Psychological Review*

Basic learning technique: *Back-propagation*

Deep Learning (DL)

Basic learning technique: **Back-propagation** – *supervised*

- *Activation* flows from input layer to output layer
- *Error* values flow from output layer to input layer

Increasing the network depth (number of hidden layers) can dramatically improve learning – given

- enough training data &
 - enough computing power
- Google trains networks with *billions* of connections
- 

*the only change
since the 1980s!*

Cognitive Science comes of age

Cognitive Science: Characterize cognition – the function of the brain – precisely

A very exciting time for
Cogsci: transitioning
to a hard science!

- ◆ critical to technology
that changes the world

Artificial Intelligence – AI

Watson (IBM) 2010
Deep Neural Networks 2010

Cognitive Science		
1950	Symbolic models (language)	①
1980	Neural models (no symbols)	②
1990	Neural models (with symbols) Optimality-based grammars	③ ④
2010	Gradient Symbolic Computing	⑤

Today's topics

- ◆ critical to an established hard science

Brain reading

2000 gross functions
2010 single objects
2020 complex events

Simple models that
classify brain activation
patterns into categories

Vegetative or locked-in?

An early diagnostic test (one patient; Owen, Adrian. 2006. *Science*)

- Ask patient inside an MRI machine personal questions (to which the scientists do not know the answer)
 - do you have a brother?
- Patient is told to respond, when hearing “Answer!”:
 - for yes, “imagine playing tennis”
 - for no, “imagine navigating your apartment”
and to stop responding when hearing “Relax!”
- Look for activity in either of two of the patient’s brain regions previously shown to associate with these two different mental activities

Results

- 5 of 6 questions: measurable response — all correct

Cognitive Science comes of age

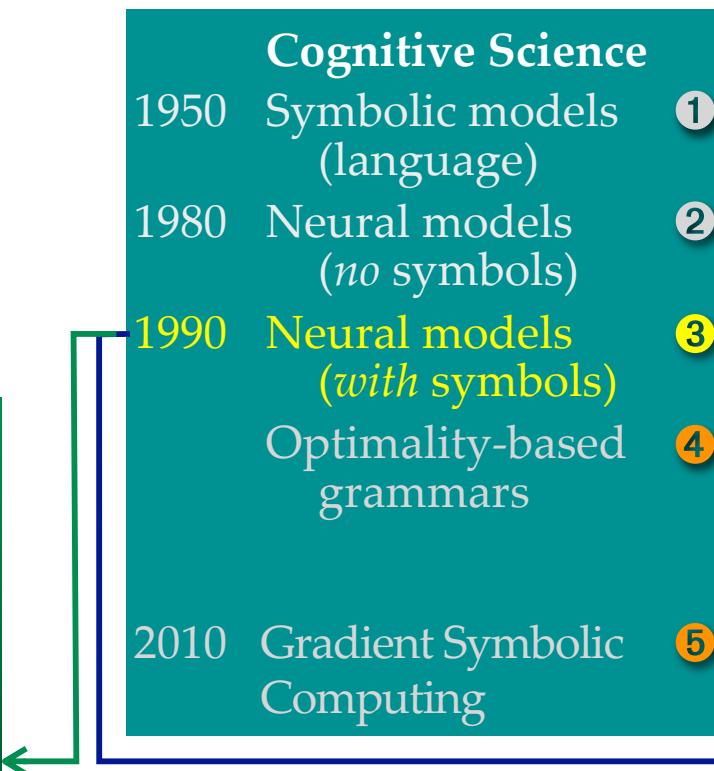
Cognitive Science: Characterize cognition – the function of the brain – precisely

A very exciting time for
Cogsci: transitioning
to a hard science!

- ◆ critical to technology
that changes the world

Artificial Intelligence – AI

Watson (IBM) 2010
Deep Neural Networks 2010



Today's topics

- ◆ critical to an established hard science

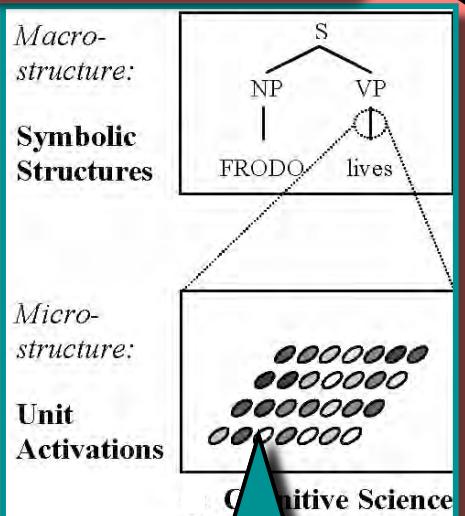
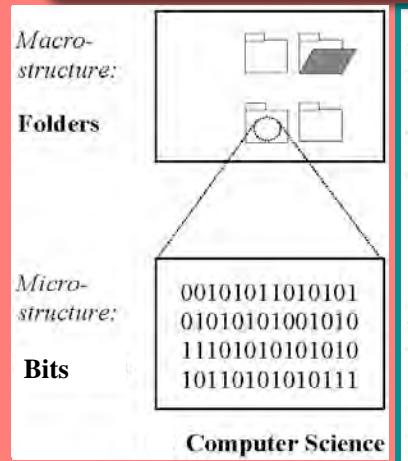
Brain reading

2000 gross functions
2010 single objects
2020 complex events

A mathematical isomorphism links levels in the human mind/brain:
high-level virtual symbolic computer,
low-level neural network computer

- ♦ Human thought can cope intelligently with an unlimited set of situations: *productivity*.
- ♦ Human knowledge allows (i) a novel situation to be broken down into familiar parts, (ii) the parts to be processed separately, & (iii) the results reassembled.

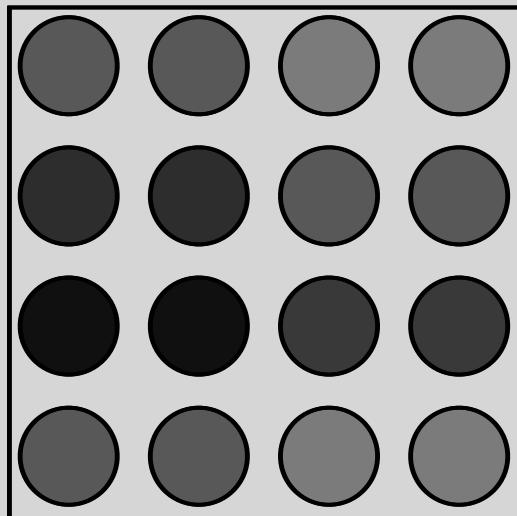
- ☞ *The combinatorial strategy* for productivity.
- ♦ The ‘parts’ can be
 - ▶ symbols arranged into structures
 - ▶ distributed patterns of neural activity.



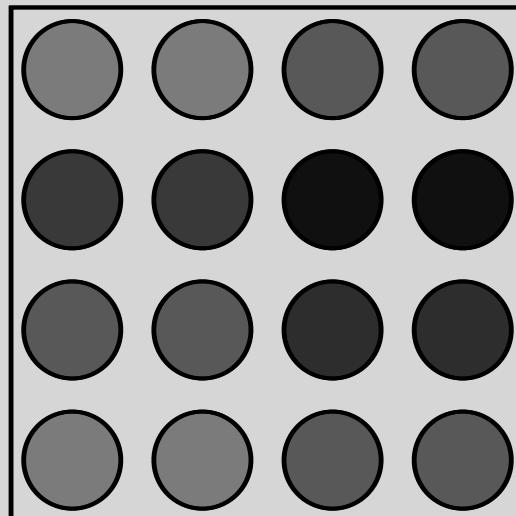
**Combinatorial
structure of
activation patterns?**

**Cognition is computation over
combinatorial structures**

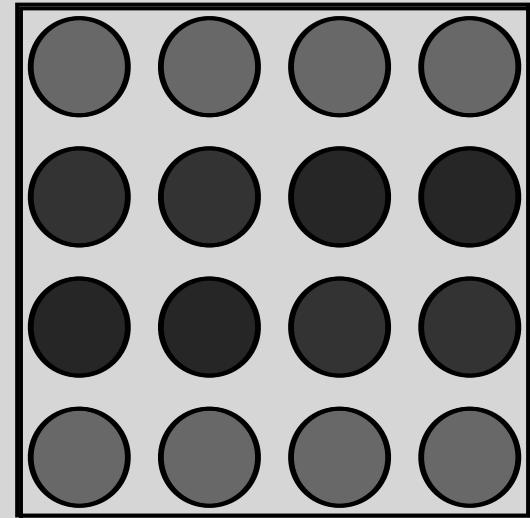
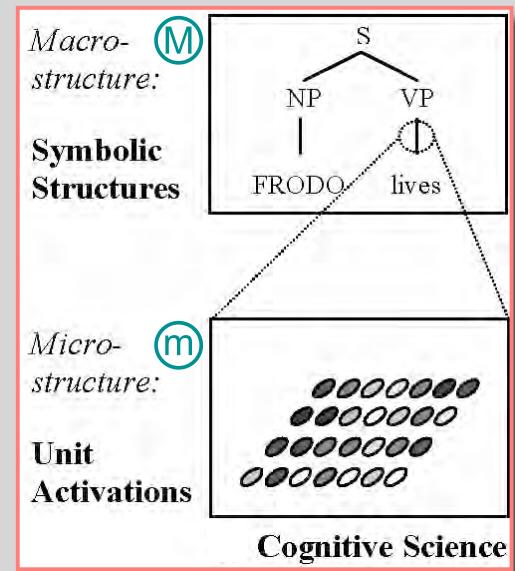
Nerd slide — Combinatorial structure in activation patterns



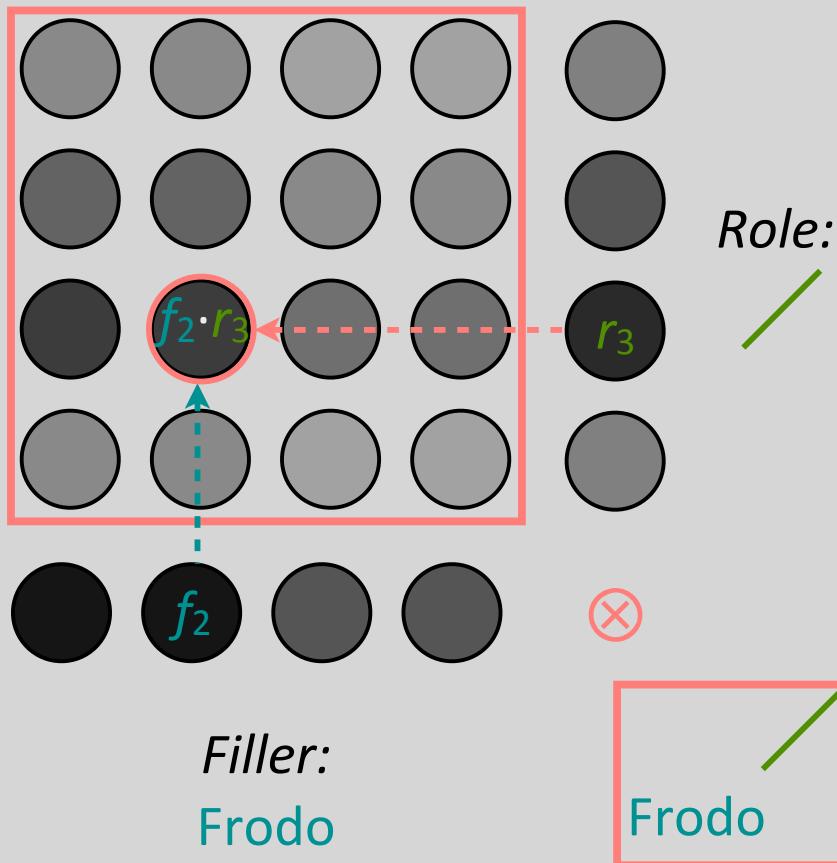
+



=

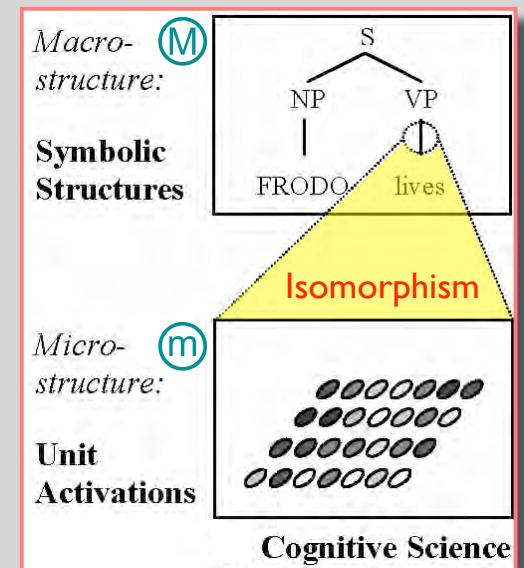


Nerd slide — Constituent construction



Filler pattern
⊗
Role pattern
=

Filler/role
constituent pattern



Tensor product
representations

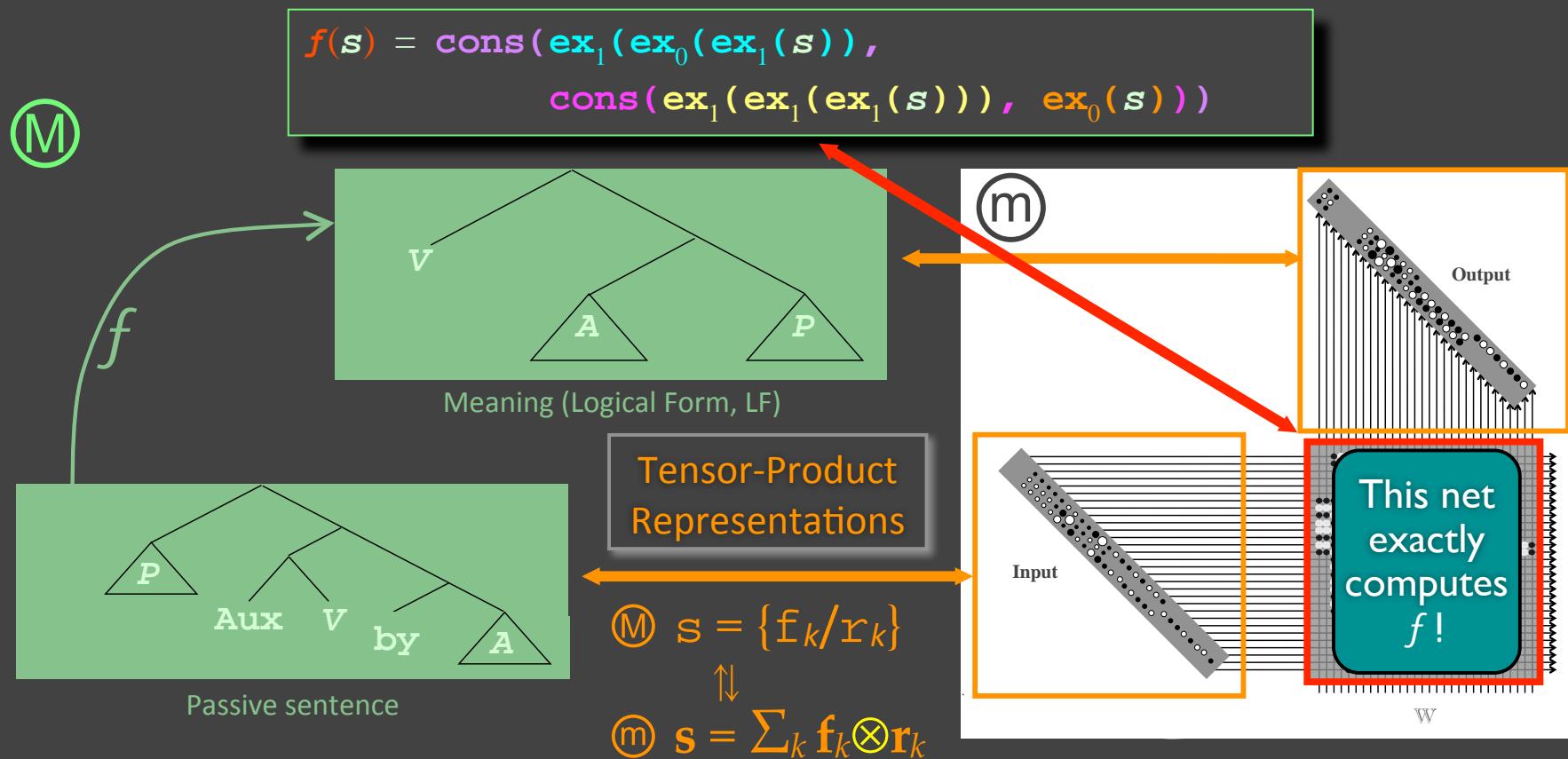
$$\textcircled{M} \quad s = \{f_k / r_k\}$$

$$\textcircled{m} \quad s = \sum_k f_k \otimes r_k$$

Nerd slide —

The isomorphism \uparrow
 Ⓜ Symbolic virtual machine
 Ⓜ Neural network

Few leaders are admired by George Bush \xrightarrow{f} admire(George Bush, few leaders)



Cognitive Science comes of age

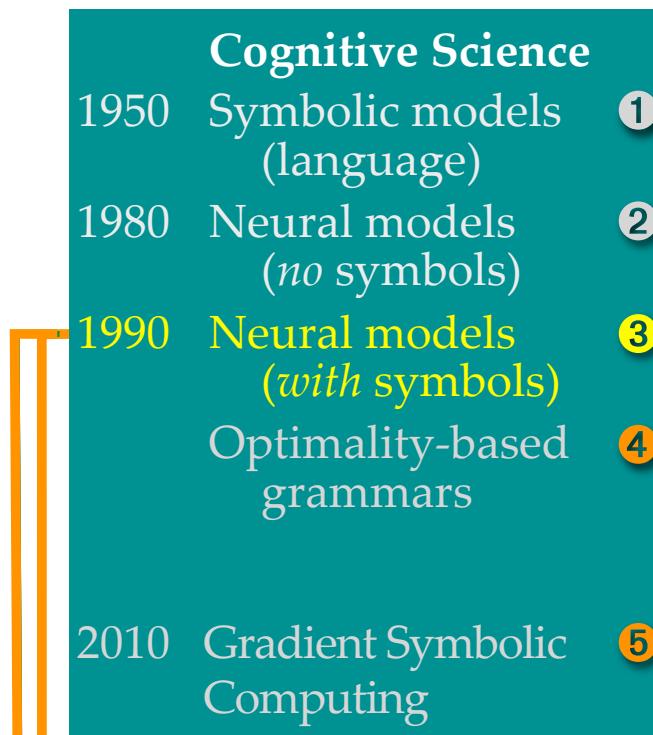
Cognitive Science: Characterize cognition – the function of the brain – precisely

A very exciting time for
Cogsci: transitioning
to a hard science!

- ◆ critical to technology
that changes the world

Artificial Intelligence — AI

Watson (IBM)	2010
Deep Neural Networks	2010
Deep Nets with symbols	2020



Today's topics

- ◆ critical to an established hard science

Brain reading

2000	gross functions
2010	single objects
2020	complex events

Cognitive Science comes of age

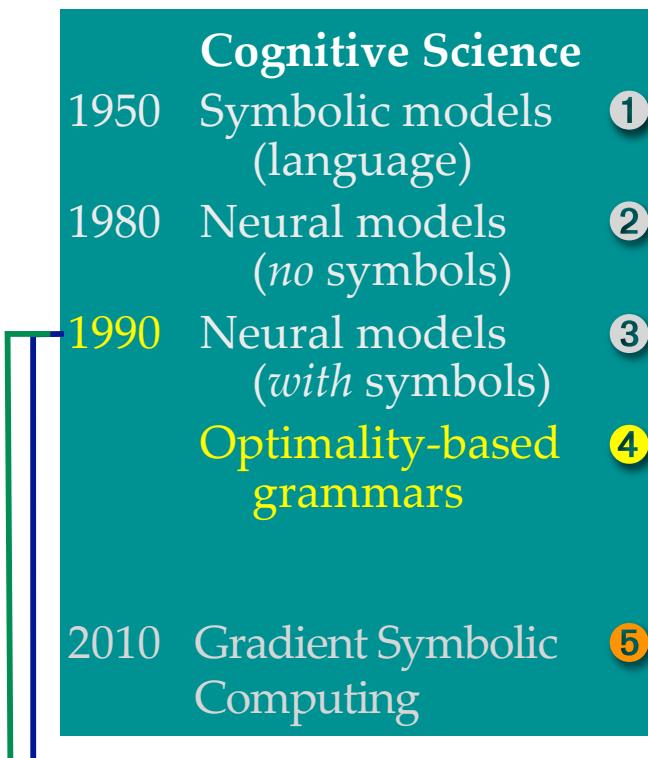
Cognitive Science: Characterize cognition – the function of the brain – precisely

A very exciting time for
Cogsci: transitioning
to a hard science!

- ◆ critical to technology
that changes the world

Artificial Intelligence — AI

Watson (IBM)	2010
Deep Neural Networks	2010
Deep Nets with symbols	2020

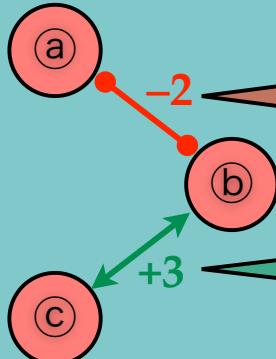


Today's topics

- ◆ critical to an established hard science

Brain reading

2000	gross functions
2010	single objects
2020	complex events



\textcircled{m} -desideratum: (a) and (b) should not both be active simultaneously (if they are \Rightarrow Harmony penalty 2)

\textcircled{m} -desideratum: (b) and (c) should both be active simultaneously (if they are \Rightarrow Harmony reward 3)

Legendre, Miyata, & Smolensky 90, 91;
Pater et al. 08 et seq.

- ◆ Computation in certain neural networks builds an activation pattern with maximal micro-Harmony (conforming maximally to the **micro-desiderata** encoded in the connection weights): ***the Hmax principle***.
- ◆ For tensor-product representations, this mathematically entails that outputs maximize macro-Harmony:
- ◆ the **macro-desiderata** in macro-Harmony define a **Harmonic Grammar**.

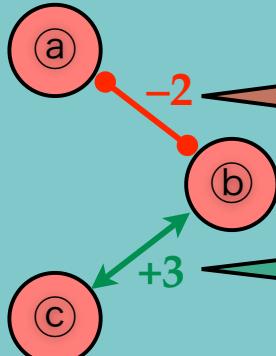
Macro desiderata for sentence structure/syntax:

SUBJECT: A sentence has a subject.

FULLINTERP: No meaningless words.

Grimshaw & Samek-Lodovici 95, 98

Computation in grammar is optimization:
knowledge = **desiderata**,
not procedure



ℳ-desideratum: (a) and (b) should not both be active simultaneously
(if they are \Rightarrow Harmony penalty 2)

ℳ-desideratum: (b) and (c) should both be active simultaneously
(if they are \Rightarrow Harmony reward 3)

- ♦ Computation in certain neural networks built on tensor products of vectors.
- If (a) & (c) both active, at (b): **Conflict!**
Impossible to satisfy both.
Need conflict-resolution: weighting.

- For tensor-product representations, this mathematically entails that outputs maximize macro-Harmony:
- the **macro-desiderata** in macro-Harmony define a **Harmonic Grammar**.

Macro desiderata for sentence structure/syntax:

SUBJECT: A sentence has a subject.

FULLINTERP: No meaningless words.

Conflict! Impossible to satisfy both.
Need a conflict-resolution method.

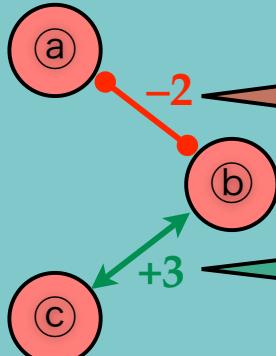
How to express <rains>?

	English			Italian		
	-3	-2	H	-3	-2	H
	SUBJECT	FULLINTERP		FULLINTERP	SUBJECT	
rains	*	!	-3	*	-2	-3
it rains		*	-2	!		-3
it rains it		* *!	-4	*!*		-6

A realistically complex case (in phonology) will be shown below

Typology by re-weighting!
Macro-desiderata, not procedure





m-desideratum : (a) and (b) should not both be active simultaneously
(if they are \Rightarrow Harmony penalty 2)

m-desideratum : (b) and (c) should both be active simultaneously
(if they are \Rightarrow Harmony reward 3)

- ♦ Computation in certain neural networks built on tensor products of vectors. If (a) & (c) both active, at (b): Conflict! Impossible to satisfy both.

If (a) & (c) both active, at (b): Conflict!
Impossible to satisfy both.
Need conflict-resolution: weighting.

- ♦ For tensor-product representations, this mathematically entails that outputs maximize macro-Harmony:
- ♦ the **macro-desiderata** in macro-Harmony define a **Harmonic Grammar**.

Macro desiderata for sentence structure/syntax:

SUBJECT: A sentence has a subject.

FULLINTERP: No meaningless words.

How to express $\langle \text{rains} \rangle$?

	English			Italian		
	-3	-2		-3	-2	
	SUBJECT	FULLINTERP	H	FULLINTERP	SUBJECT	H
rains	*	!	-3	*	-2	-3
it rains		*	-2	*		-2
it rains it		*	-4	*		-4

Typology by re-weighting!
not procedure

Weaker constraint is not
'turned off' as in other theories

A realistically complex case (in phonology) will be shown below

OT is a **general formalism for grammar**; applied in

- ◆ phonetics, phonology, morphology, syntax, semantics, pragmatics, discourse
- ◆ acquisition, language change, bilingualism, code-switching

Many computational theorems and software for computing outputs, typologies; learning
(actually used by working linguists)

Often enables more insightful or explanatory analysis (not necessarily *simpler*)

◆ the macro-desiderata in macro-Harmony define a **Harmonic Grammar**. *Key variant:*

◆ **Optimality Theory** (OT)

- ▶ macro-desiderata = *constraints* universal: the same in all languages
- ▶ constraints are ranked not weighted
Engl: SUBJECT ≫ FULLINTERP; Ital: reverse

☞ can formally compute the typology of possible human grammars from all rankings of the universal constraints

Computation in grammar is optimization:
knowledge = **desiderata**, not procedure

⇒ *Radical new theory of grammar*

OT is a **general formalism for grammar**; applied in

- ❖ phonetics, phonology, morphology, syntax, semantics, pragmatics, discourse
- ❖ acquisition, language change, bilingualism, code-switching

Many computational theorems and software for computing outputs, typologies; learning (actually used by working linguists)

Often enables more insightful or explanatory analysis (not necessarily *simpler*)

OT gives new and formally precise answers to the most fundamental questions of grammatical theory:

What, exactly, is it that all human grammars share?

The constraints:

desiderata for combinatorial structure

How, exactly, may human grammars differ?

Only in the ranking (or weighting) of the constraints.

❖ the macro-desiderata in macro-Harmony define a **Harmonic Grammar**. *Key variant:*

❖ **Optimality Theory** (OT)

- macro-desiderata = *constraints* universal: the same in all languages
- constraints are ranked not weighted
Engl: SUBJECT > FULLINTERP; Ital: reverse

☞ can formally compute the typology of possible human grammars from all rankings of the universal constraints

Computation in grammar is optimization: knowledge = **desiderata**, not procedure

⇒ Radical new theory of grammar

From sequential symbol-manipulating procedures to achieving (simultaneous) desiderata: optimization (a real example)

From sequential symbol-rewriting to Harmony maximization

Lardil How to say 'wooden axe'?

If object:

muŋkumuŋku = muŋkumuŋku-n
stem-suffix

If subject:

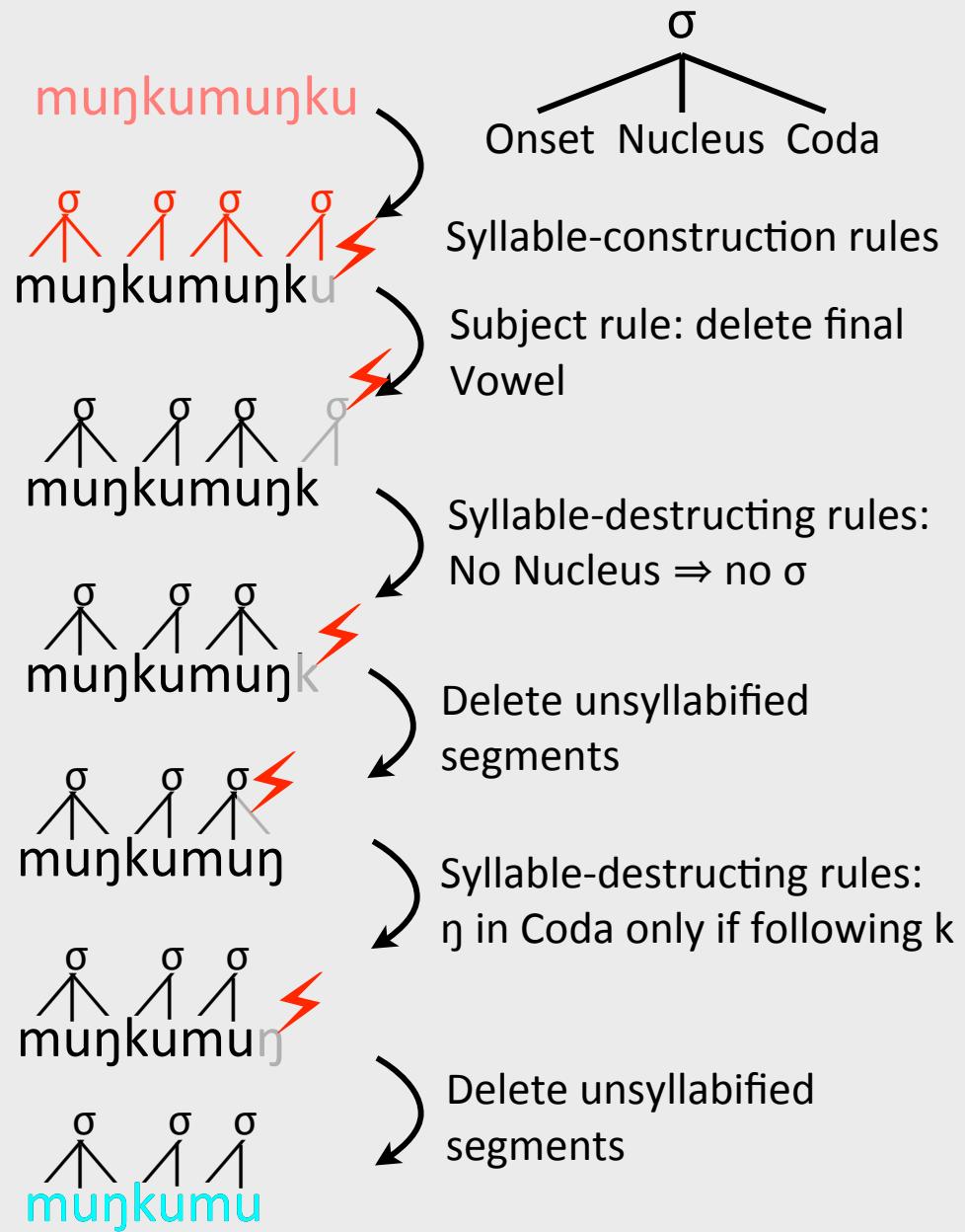
muŋkumu

From what general knowledge
can the speaker compute this,
and all the other forms?

A grammar:

What form of computation?

And now for something
completely different ...



à la Wilkinson 1983

From sequential symbol-rewriting to Harmony maximization

<i>Stem</i>	<i>Nominative</i>	{undominated}	*INSERT(V)	NOM	ALIGN	*INSERT(C)	*DELETE
muŋkumunuŋku	'wooden axe'						
	muŋ.ku.muŋ.ku			* !			
	muŋ.ku.muŋk	*COMPLEX ! ...			*		*
	muŋ.ku.muŋ	*CODACOND !			*		**
	muŋ.ku.mu.ŋa		* !				
👉	muŋ.ku.mu				(*)		(*) (*) (*)
	muŋ.ku				*		*** * ! *

wite	'inside'						
👉	wi.te			(*)			
	wit	*MINWD !			*		*

řil	'neck'						
	řil	*MINWD !					
	řil.a	*ONSET !	*				*
	ři.la		*		* !		
👉	řil.ta		(*)			(*)	
	řil.tat		*			* * !	

From sequential symbol-rewriting to Harmony maximization

Stem	Nominative	{undominated}	*INSERT(V)	NOM	ALIGN	*INSERT(C)	*DELETE
muŋkumunŋku	'wooden axe'						
	muŋ.ku.muŋ.ku			*	!		
	muŋ.ku.muŋk	*COMPLEX ! ...			*		*
	muŋ.ku.muŋ	*CODACOND !			*		**
	muŋ.ku.muŋa		*	!	← Fatal here [there's a better option]		
☞	muŋ.ku.mu					⊗ ⊗ ⊗	
	muŋ.ku	Viable constraints			*		* * * * ! *
wite	'inside'						
☞	wi.te				⊗		
	wit	*MINWD !			*		*
řil	'neck'						
	řil	*MINWD !					
	řil.a	*ONSET !		*			*
	ři.la			*		*	
☞	řil.ta			⊗	← Tolerated here [there's no better option]		
	řil.tat		*				

Cognitive Science comes of age

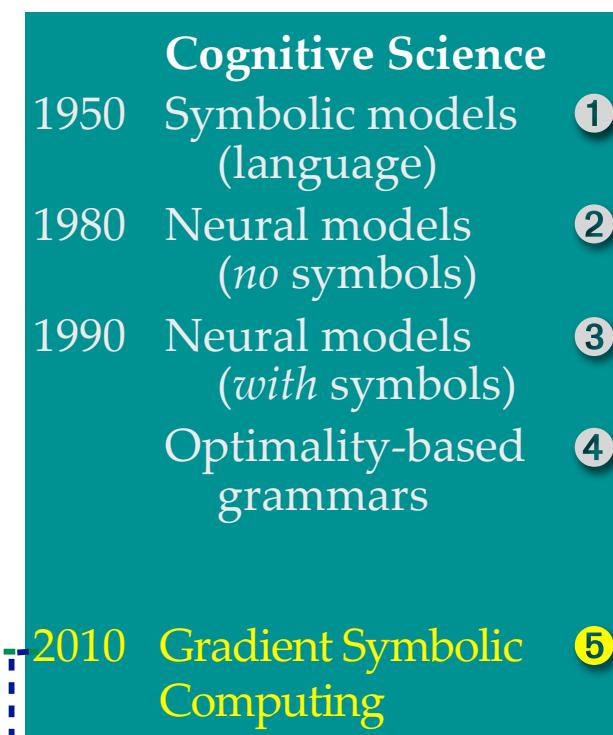
Cognitive Science: Characterize cognition – the function of the brain – precisely

A very exciting time for
Cogsci: transitioning
to a hard science!

- ◆ critical to technology
that changes the world

Artificial Intelligence — AI

Watson (IBM)	2010
Deep Neural Networks	2010
Deep Nets with symbols	2020



Today's topics

- critical to an established hard science

Brain reading

2000	gross functions
2010	single objects
2020	complex events

♦ Example: Gradient input, discrete output

♦ French liaison

petit ami vs. petit copain vs. petite copine
 pe.ti.ta.mi pe.ti_.co.pain pe.tit.co.pine

$$\text{peti}(0.5*t) + (0.3*t + 0.3*n + 0.3*z)\text{ami} \quad \begin{matrix} \text{input} \\ \downarrow \\ \text{pe.ti.ta.mi} \end{matrix} \quad \begin{matrix} \text{optimize} \\ \text{output} \end{matrix}$$

$$\text{peti}(0.5*t) + \text{copain} \quad \text{peti}(1*t) + \text{copine} \quad \begin{matrix} \downarrow \\ \text{pe.ti.co.pain} \end{matrix} \quad \begin{matrix} \downarrow \\ \text{pe.tit.co.pine} \end{matrix}$$

♦ Gradient Symbol Structures: Symbols in structures have *continuous activation values*.

$$\begin{matrix} & & \\ & 0.7A & 0.4A \\ & + 0.2B & + 0.9C \end{matrix}$$

A in role blend: $0.7r_{\text{left}} + 0.4r_{\text{right}}$

Left child role filled by blend of symbols

GSC: novel continuous neural computation yields (discrete or gradient) optimal abstract combinatorial outputs.

⇒ Radical new theory of how abstract information is represented in the mind

Cognitive Science comes of age

Cognitive Science: Characterize cognition – the function of the brain – precisely

A very exciting time for
Cogsci: transitioning
to a hard science!

- ◆ critical to technology that changes the world

Artificial Intelligence — AI

Watson (IBM)

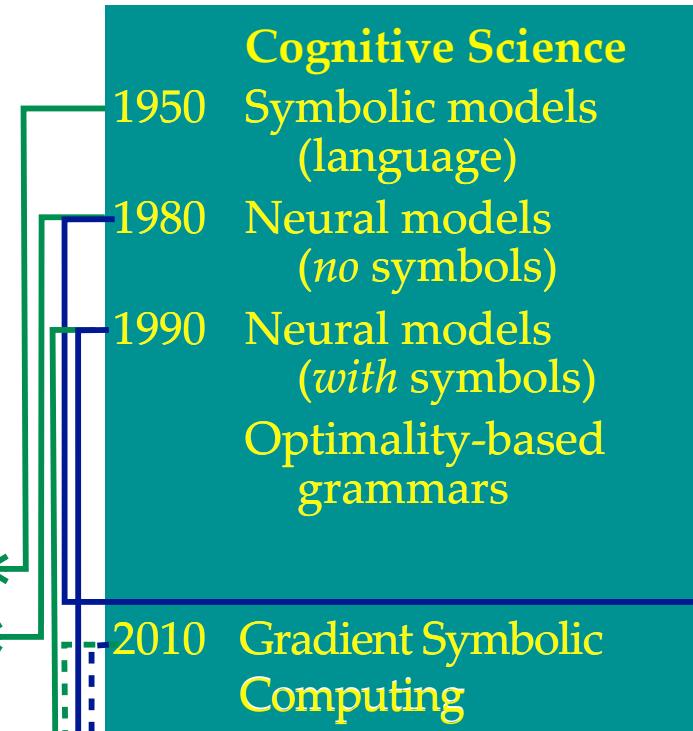
2010

Deep Neural Networks

2010

Deep Nets with symbols

2020



- ◆ critical to an established hard science

Brain reading

2000 gross functions
2010 single objects
2020 complex events

Cognitive Science comes of age

Cognitive Science: Computational approaches to mind and brain precisely

A very exciting time for
Cogsci: transitioning
to a hard science!

- ◆ critical to technology
that changes the world

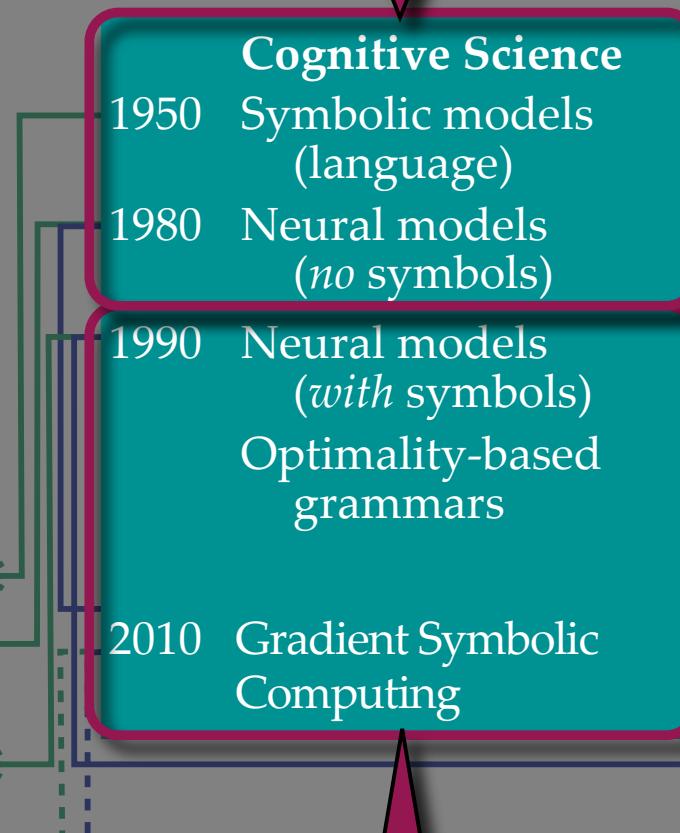
Artificial Intelligence — AI

Watson

Deep Neural Networks

2010

2010



Today's topics

- ◆ critical to an established hard science

Brain reading

2000 gross functions
2010 single objects
2020 complex events

How language is profoundly shaped by its neural substrate

That's all folks!

Thank you for your attention.