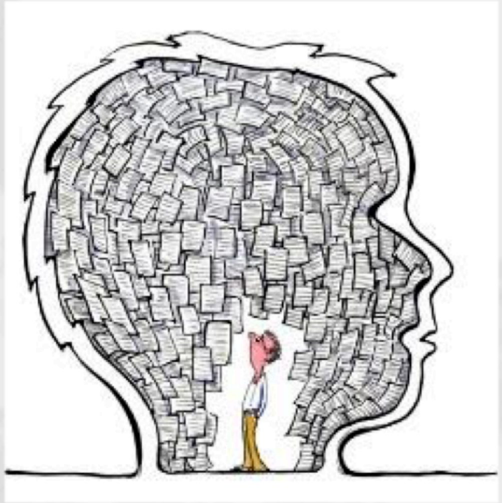
from memory processes to lexical self-organisation: a biologically-motivated integrative view of the morphological lexicon

VITO PIRRELLI - Comphys Lab
INSTITUTE FOR COMPUTATIONAL LINGUISTICS, PISA CNR ITALY

a premise: words are…

stored representations or dynamic processes?

- computationally (as well as psychologically) words prove to be elusive theoretical constructs, retaining features of **both** stored **representations** and dynamic **processes**



words are…

- “permanent”
  - … but their level of resting activation changes over time and contexts
- “stored” word-wise
  - … but can be “perceived” morpheme-wise
- “accessed/retrieved”
  - … but can be produced “on-line”
- associatively related
  - … but can “compete” with one another for activation primacy and selection
- exhibiting degrees of wordlikeness
  - … modulated through a wide range of frequency effects

the role of frequency

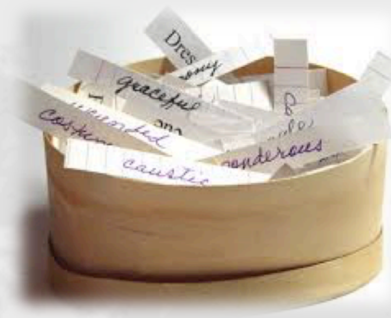
- token frequency of an inflected form facilitates lexical access and correlates negatively with response latencies in visual lexical decision (Taft and Forster, 1975; Whaley, 1978)

- the more frequent an inflected form is relative to its base (e.g. walked vs. walk), the more salient the whole is relative to its parts (Hay and Baayen, 2005)

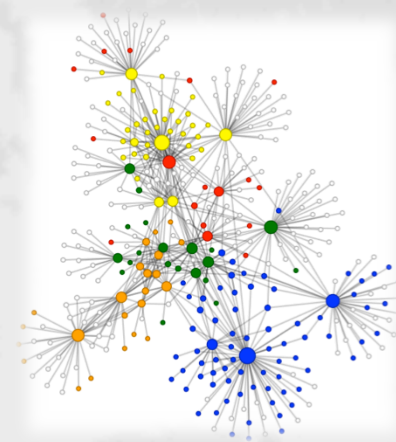
- a more uniform frequency distribution over members of the same inflectional paradigm makes them more readily accessible (Moscoso del Prado Martín et al., 2004; Baayen et al., 2006), favouring a better allocation of memory resources

is there a “place” for words?

- in traditional wisdom, word knowledge is thought to reside in the **mental lexicon**, a kind of brain dictionary that contains information regarding words’ representational features, but …



- a more dynamic view is possible: words are **stimuli** and they cause a particular change in the **activation** state of the brain, for example:
  - an association with a particular concept
  - an expectation for another word to come in a sentence
  - an association with a class of possible lexical competitors



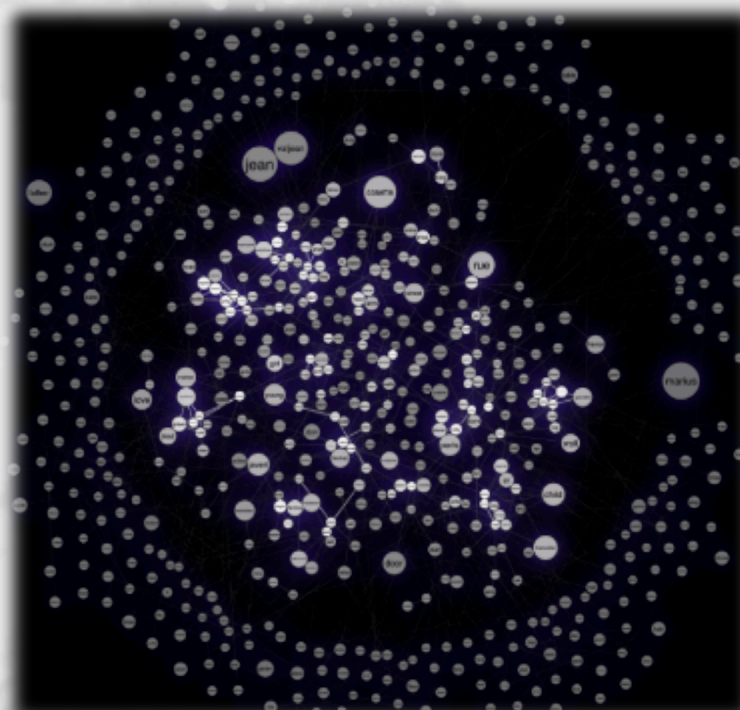
- neuro-functional evidence tells us that words are not localised in a single brain region but are themselves **emergent properties** of the functional interaction between different brain regions

two opposing camps …

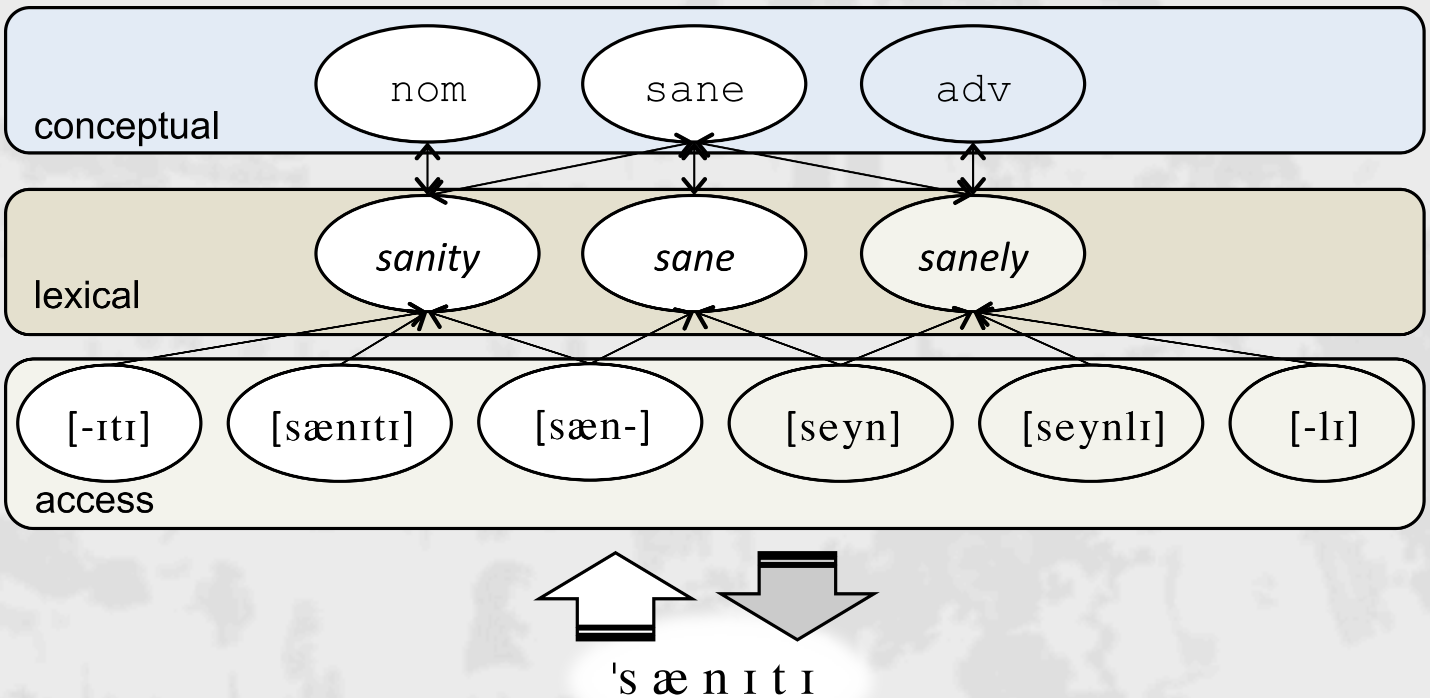
structured
(representational or memory-based)

- mach-t
- ge-mach-t
- ge-frag-t
- k-a-t-a-b-a
- ya-kt-u-b-u
- book
- hand-book
- de-rid-ere
- rid-iamo
- telefon-iamo

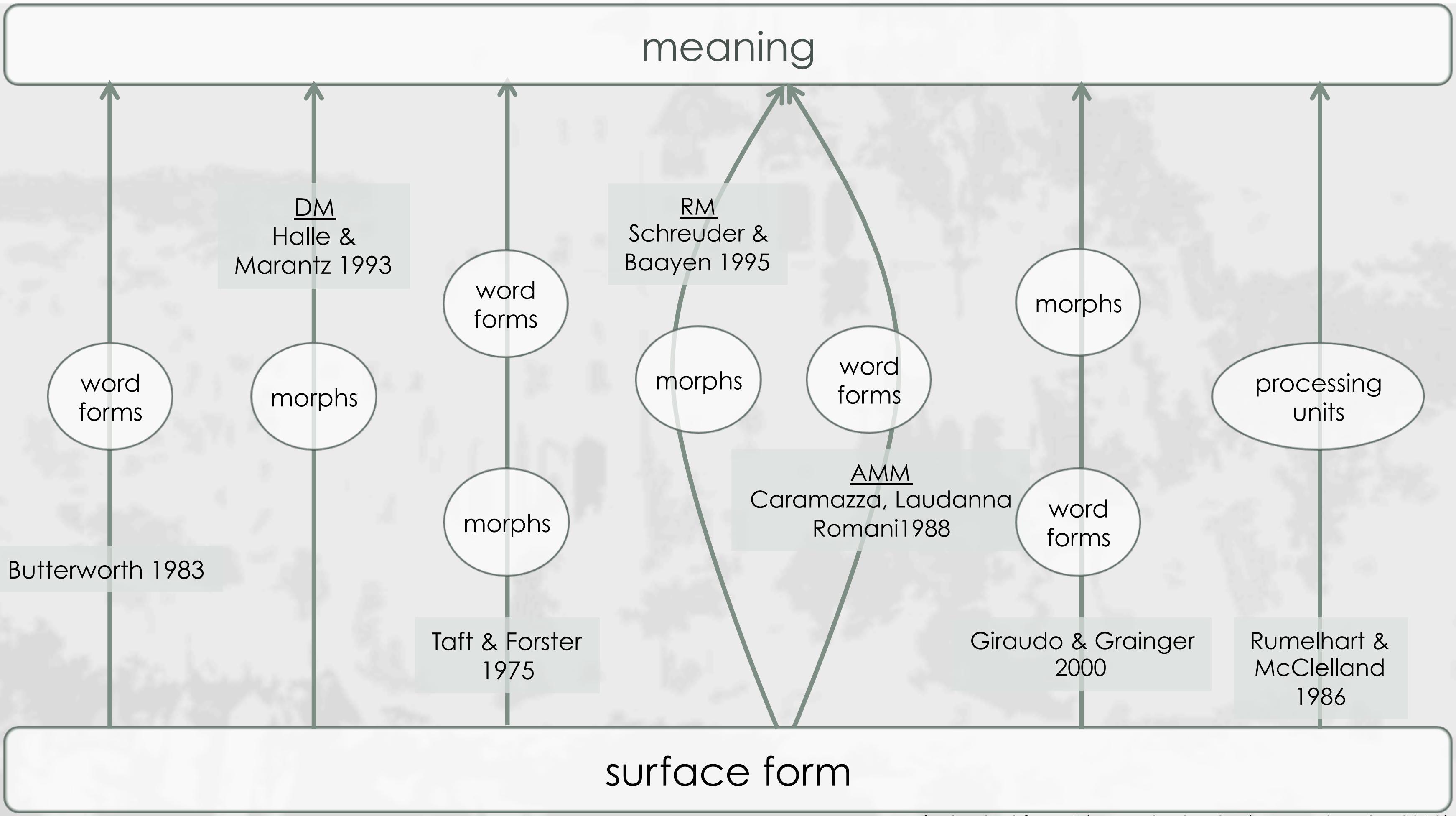
unstructured
(epiphenomenal or process-based)



multiple race lexical access



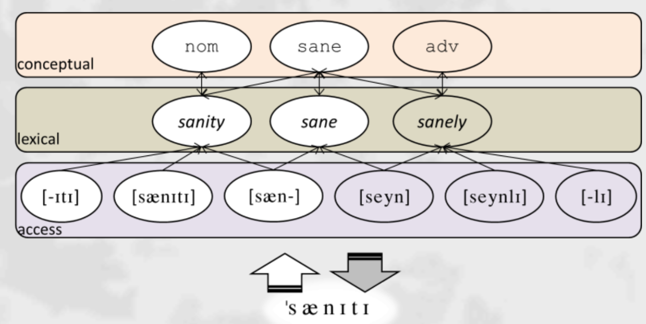
lexical architectures



(adapted from Diependaele, Grainger & Sandra 2012)

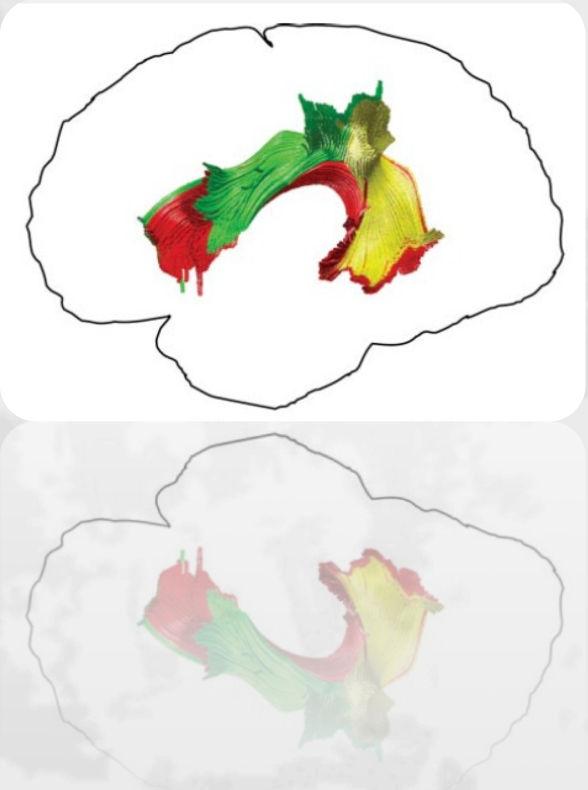
interim balance

- any cognitively-motivated hypothesis of lexical architecture must assume that accessing a word leaves its traces in the lexicon
- accessing an item must have two consequences:
  - modify the item's representation
  - increase the probability that the item will be successfully processed in the future
- many current models assume that access representations are already in place, somewhat given, internalised objects
  - principled distinction between lexical representations on the one hand, and processes applying to representations on the other hand
- these models are “distinctive”, in that they draw a sharp boundary between memory and processing



towards an "integrative" view

- lexical representations are acquired dynamically
- little is understood in modelling lexical storage and access if we do not explain how lexical representations come into existence in the first place
- words do not define an independently-given content, but are <u>input stimuli</u> causing a particular change in the activation state of the lexicon (memory traces)
- memory traces are **both** representational units (i.e. the specialised, long-term activation patterns indexing individual input stimuli in the mental lexicon), **and** processing units (dynamically responding to particular classes of stimuli)

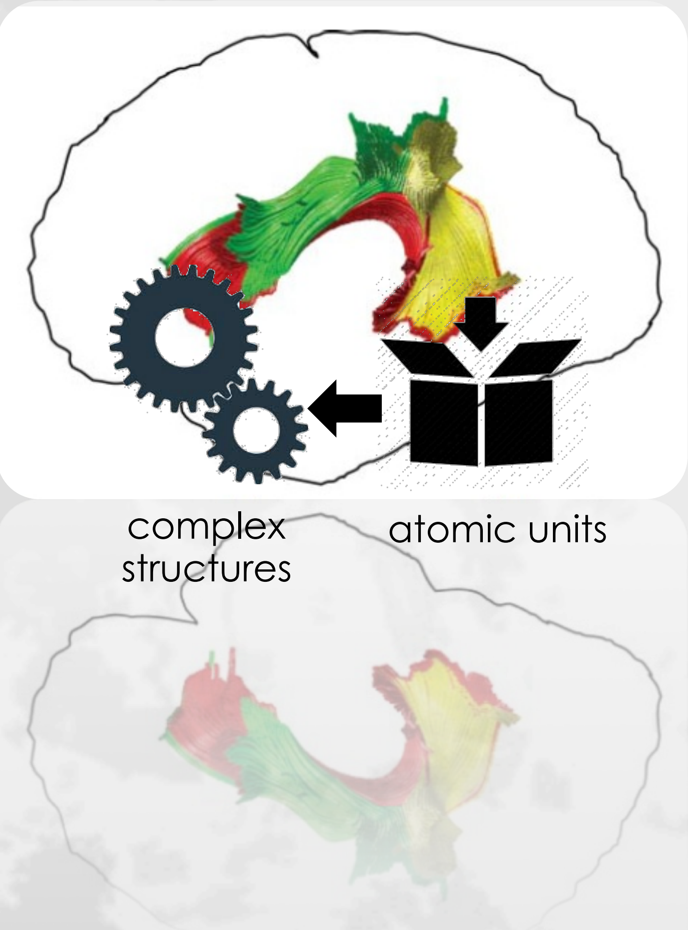


neuro-functional implications

- the "correspondence hypothesis" (Miller & Chomsky 1963, Clahsen 2006)

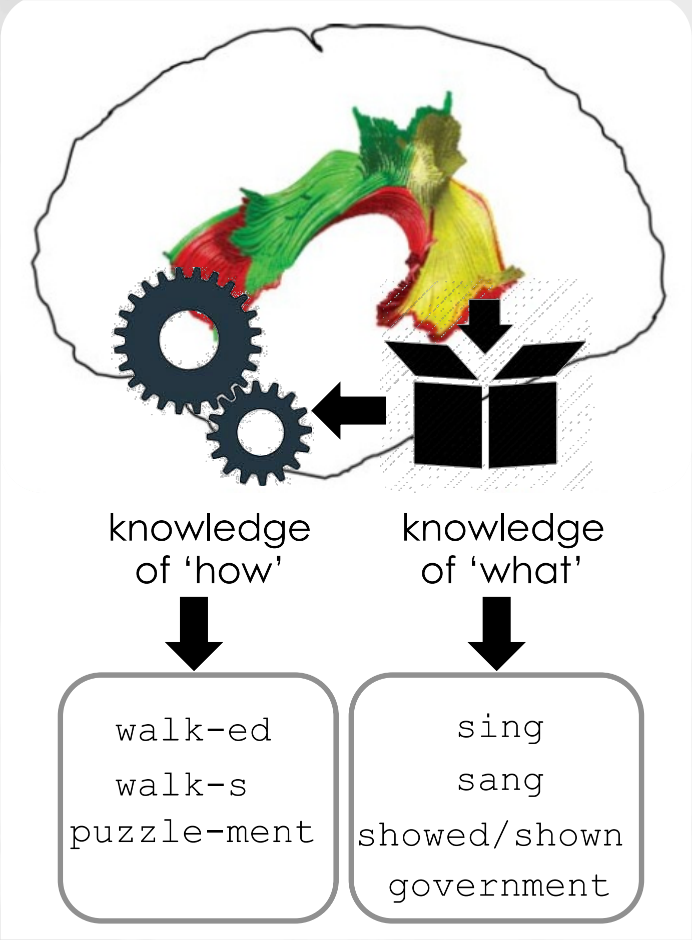
  "rules and principles of grammar organization are <u>directly</u> mirrored by the mental processes and neural structures whereby speakers understand and produce language"

  - declarative memory = mental lexicon
  - procedural memory = rule system



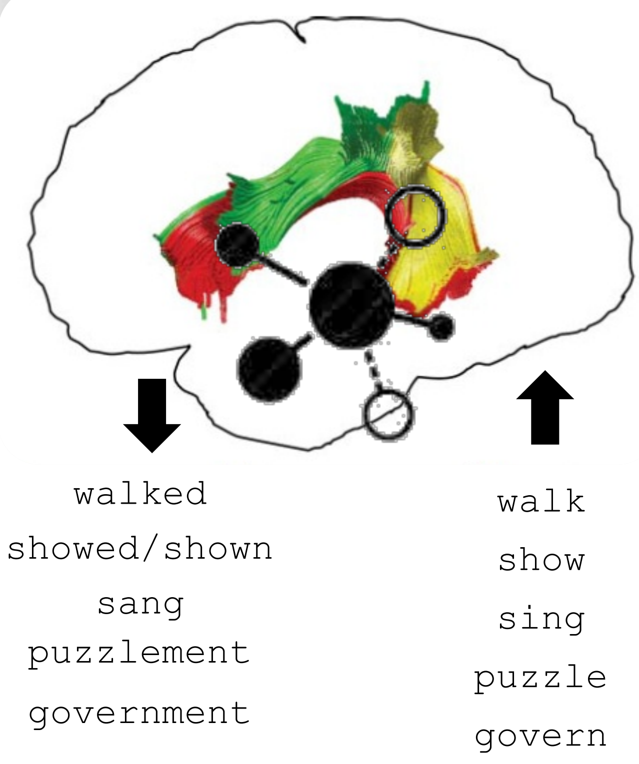
the dual route "D-P" model

- Prasada & Pinker (1993), Ullman (2001), Pinker & Ullman (2002)
- lexicon (associative patterns)
  - lexical bases
  - affixes
  - non-affixed morphologically-complex words (irregulars)
  - doublets
  - high-frequency words
- rules (symbol processing)
  - affix-based default forms (regulars)
- modularity
  - partially non-overlapping mechanisms
  - dissociation regular vs. irregular effects
- domain generality
  - stored forms pattern with known facts/events
  - computed forms pattern with acquired skills /habits
- brain localization
  - prefrontal-basal ganglia
  - temporo-parietal



connectionism

- Rumelhart & McClelland 1986, Bates & MacWhinney 1989, Elman et al. 1996, Bybee 1995
- all lexical and grammatical knowledge is learned, represented and computed over a unique associative memory
- no categorical distinction between compositional (regular) and noncompositional (irregular) forms
- non modularity
  - single associative mechanism
  - no dissociation effects predicted
- domain generality
  - brain structures subserve nonlinguistic as well as linguistic processes, but may contain domain-specific circuits
- left hemisphere distributed localization



an interim balance

DUALISM

- the idea that default rules develop in an all-or-nothing fashion, independently of exceptions and apply in a context-INsensitive way is not supported by a broadening range of empirical evidence
- frequency effects reverberate on all levels of lexical organisation and it is impossible to capture them through a redundancy-free lexicon

CONNECTIONISM

evidence of selective involvement of brain areas functionally specialised for language processing, control and storage does not lend support to the connectionist hypothesis of a holistic undifferentiated network of processing units

- at our current level of understanding, it is very difficult to establish a **direct correspondence** between language-related categories and macro-functions (rules vs. exceptions, grammar vs. lexicon) on the one hand, and neurophysiological correlates on the other hand
- as an alternative approach to the problem, we could focus on an bottom-up investigation of **basic** neurocognitive functions (e.g., serial perception, storage and alignment) to assess their involvement in language processing, according to an **indirect correspondence hypothesis**

indirect correspondence

- core processing functions:
  - (co)activation
  - binding
  - integration
  - maintenance
  - reverberation
  - storage
  - access/recall

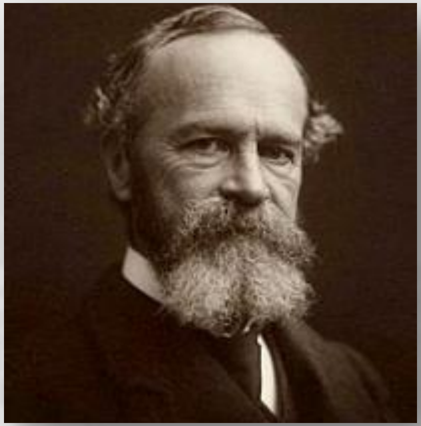
- higher-level functions:
  - serial recoding
  - lexical acquisition
  - emergent linguistic structure
  - generalisation
  - prediction
  - composition

by investigating the interaction of core processing functions and their neuroanatomical correlates we hope to shed light on higher-level functions and principles of lexical processing, and understand their role in language

correlative learning

correlative learning

- "… when two elementary brain-processes have been active together or in immediate succession, one of them, on re-occurring, tends to propagate its **excitement** into the other" (William James, 1890)



- correlation as the basis of:
  - synaptic plasticity (Hebbian rules)
  - learning and memory
  - association
  - co-activation/competition of processing units

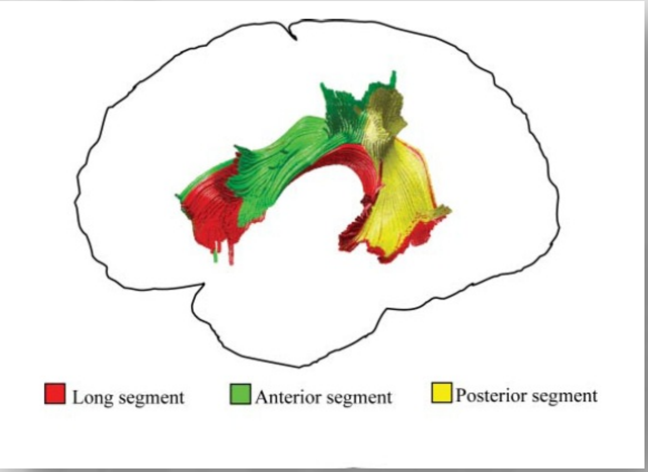
- **CL** provides a psycho-computational **framework** bringing the **dualism** between representations and processes to underlying **unity**

correlative learning and the brain

- learning is a process that generates a brain that is different from the brain prior to learning

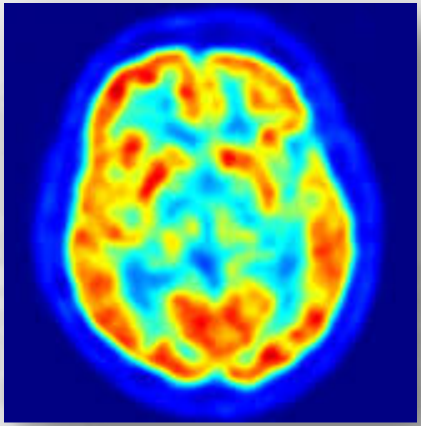
- the results of learning are **memories**

- memories are laid down in spatial patterns of synaptic connectivity making neural assemblies fire either in synch (co-activation) or sequentially (time-bound chains)



correlative learning

- activation-based processing
  - processing an input stimulus consists in competitive activation of neurons firing in synch (**neural assemblies**)
- time correlation (firing chain)
  - a time-series of stimuli produces a **chain** of consecutively activated neural assemblies
- specialisation
  - the more often an assembly fires the more it is likely to fire again (by strengthening connections to input)
  - the more often a chain of assemblies fires, the more routinized the chain will get (by strengthening connections between assemblies firing in immediate succession)



word processing and word storage

- according to this view, and contrary to both representational and epiphenomenal models of word memories, words are **memorized** as cached **assembly chains** (processing responses)

- **storage** thus depends on **processing**, as it consists in routinized assembly chains

- in turn, **processing** is **memory-based** and consists in the short-term reactivation of an assembly chain successfully responding to the input word in the past

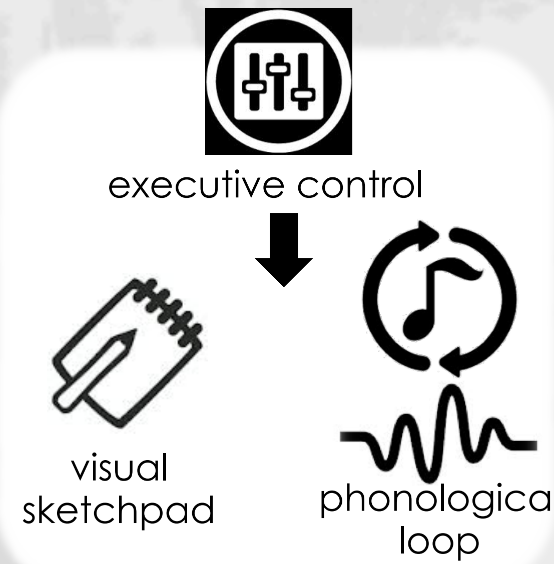
functional correlates of memory in the brain

Working Memory (WM)

- **WM** refers to the temporary retention of information that was just experienced or just retrieved from long-term memory but no longer exists in the external environment
- these representations are short-lived, but can be stored for longer periods of time through **active maintenance** or **rehearsal** strategies
- a network of brain regions, including the **prefrontal cortex** (PFC), is critical for the active maintenance of internal representations that are necessary for goal-directed behaviour
- thus, WM is not localized to a single brain region but is itself an **emergent property** of the functional interactions between the PFC and the rest of the brain

from Baddeley's WM model ...

- **strength**: not a simple container but an integrated multi-functional system involving a short-term buffer, a rehearsal mechanism (based on sub-vocal articulation) and executive control
- **weakness**: difficult to integrate long-term and short-term memory effects to account for "memory chunking" and the beneficial effects of familiar sequences on their short-term retention, under the interpretation of a sharp separation of short-term and long-term memory structures

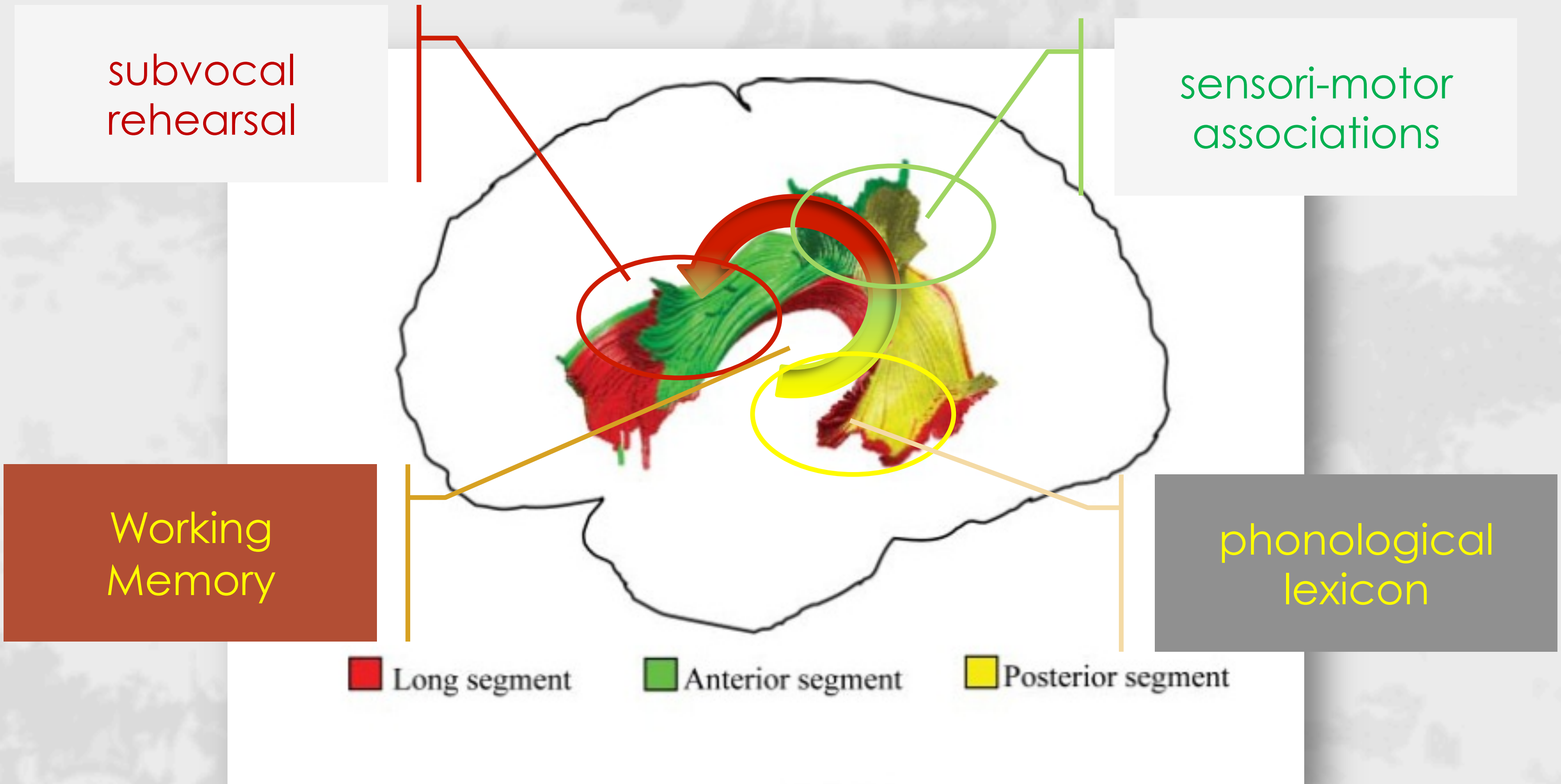


is this entirely new?

- Wernicke thought that **paraphasia** was related to the loss of a higher internal monitoring function which relied on intact connections between Wernicke's and Broca's areas:

"the unconscious, repeated activation and simultaneous mental reverberation of the acoustic image which exercises a continuous monitoring of the motor images" (Carl Wernicke, 1874)

… to WM as a dynamic system

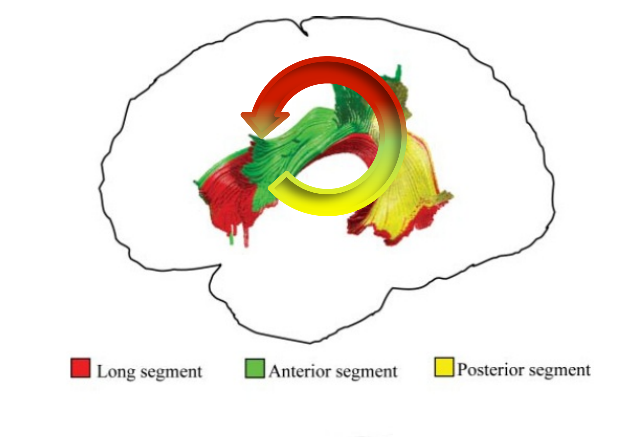


HICKOK, G. M., POEPPEL, D., 2004. Dorsal and ventral streams: a framework for understanding aspects of the functional anatomy of language. Cognition, 92: 67-99.

D'ESPOSITO, M., 2007. From cognitive to neural models of working memory. Philosophical Transactions of the Royal Society B: Biological Sciences, 362:761-772.

functional properties

- an input stimulus **activates** a neural circuit for a short time (from one to a few seconds)
- activated circuits are sustained through **reverberatory** mechanisms in the perisylvian network
- reverberation allows for **integration** of circuits in LTM
- LTM structured circuits develop as the LTM response to recurrent time-series stimuli
- by alleviating the work burden on reverberatory mechanisms, LTM structured stimuli **augment** the STM capacity of retaining longer time-series of stimuli
- in line with a new conception of WM as a limited resource of attentional capacities flexibly **distributed** among items maintained in memory

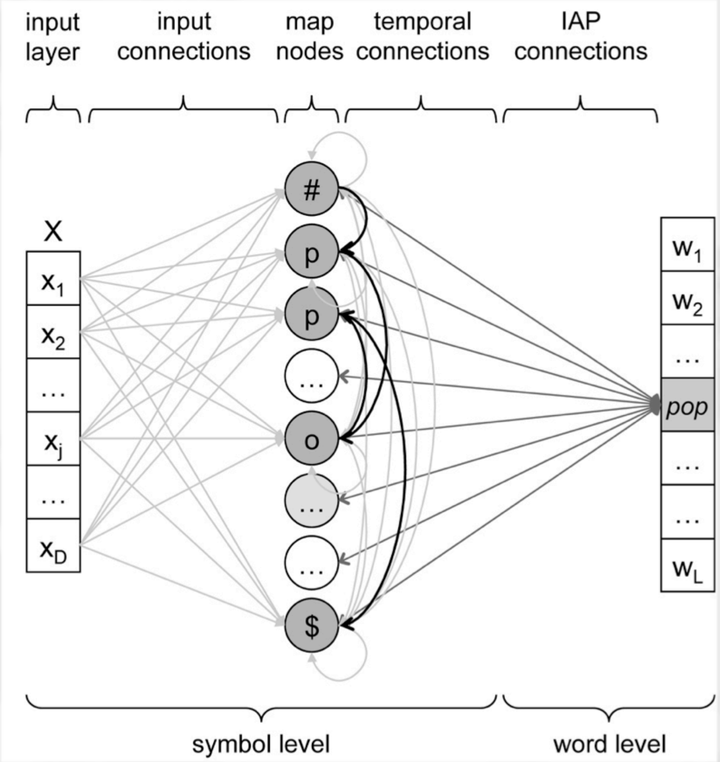


WEI JI MA, MASUD HUSAIN & PAUL M BAYS, 2014. Changing concepts of working memory. Nature Neuroscience, 17 (3): 347-356.

correlative learning, memory & brain maps

temporal brain maps

- spatially layered memory nodes learn to selectively fire upon seeing an individual symbol in a specific time frame
- two levels of connectivity
  - "input" connections from each node to input layer
  - re-entrant "temporal" connections from each node to any other node
- words are input as **time series** of symbols, by showing the map one symbol at a time



temporal brain maps (II)

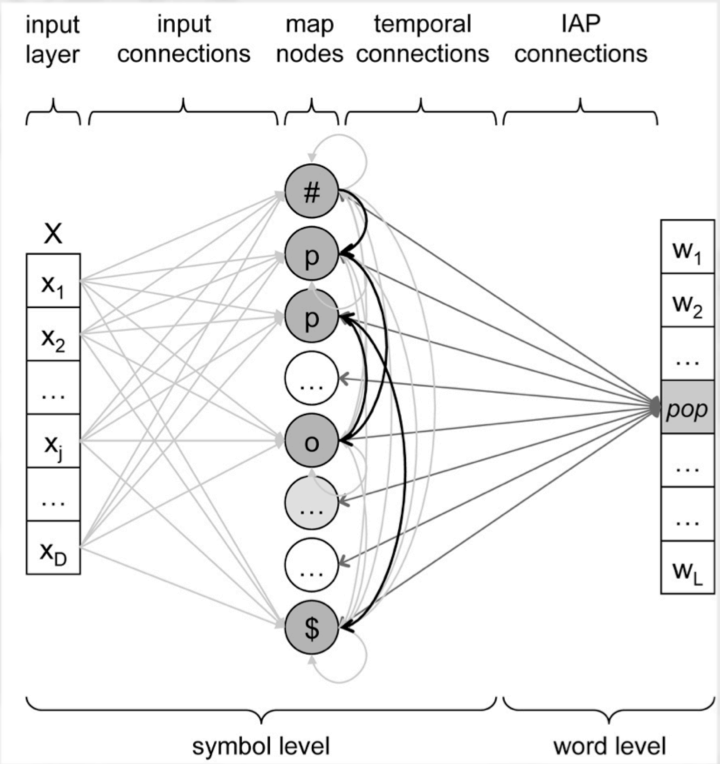
- upon being shown a symbol at time *t*, map nodes activate **concurrently** and **compete** for activation primacy

$$y_i(t) = \alpha \cdot y_{S,i}(t) + (1-\alpha) \cdot y_{T,i}(t)$$

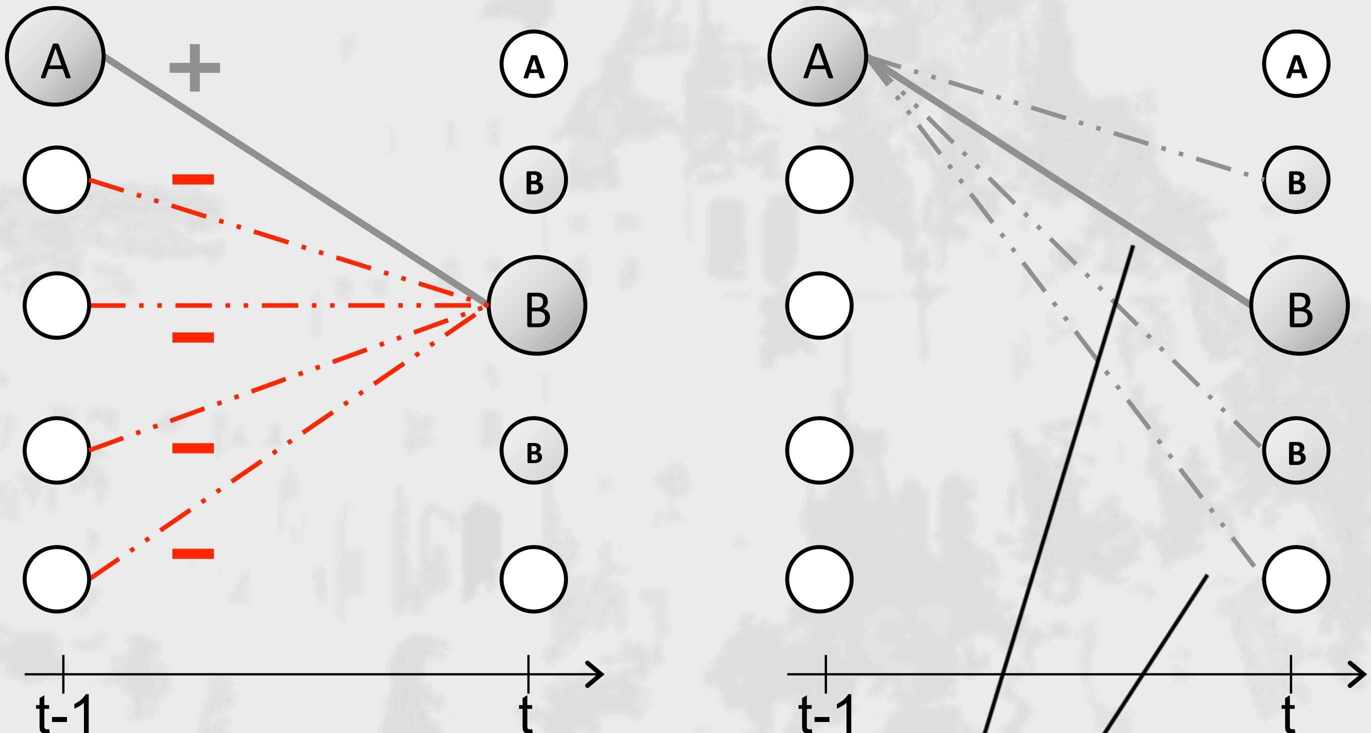
$$y_{S,i}(t) = \sqrt{D} - \sqrt{\sum_{j=1}^{D}[x_j(t) - w_{i,j}(t)]^2}$$

$$y_{T,i}(t) = \sum_{h=1}^{N}[y_h(t-1) \cdot m_{i,h}(t)]$$

- the winning node (or **BMU**) and its neighbours get a prize
  - "what" connections are **potentiated**
  - "when" connections to the BMU at time *t-1* are **potentiated**
  - "when" connections of losing nodes are **depressed**



Hebbian learning



low pointwise entropy

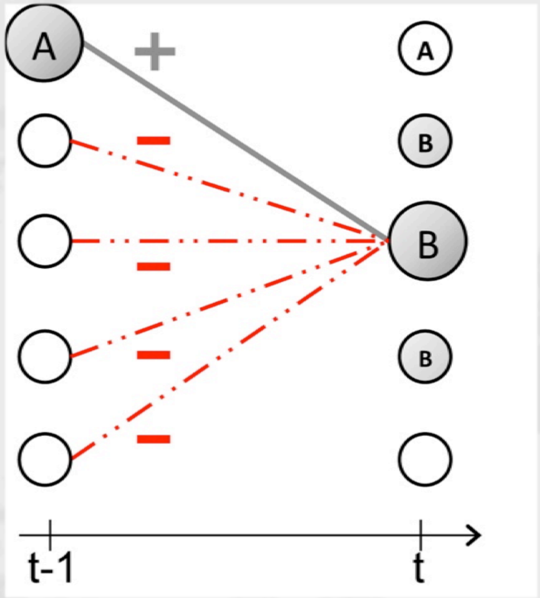
high pointwise entropy

temporal brain maps (III)

- correlative equations are strongly reminiscent of **Rescorla-Wagner** equations

$$\Delta m_{i,h}(t) = \gamma_T(t_E)\cdot c_{T,i}(t)\cdot[1 - m_{i,h}(t) + \beta_T(t_E)] \quad for \;\; h = BMU(t-1)$$

$$\Delta m_{i,h}(t) = \gamma_T(t_E)\cdot c_{T,i}(t)\cdot[0 - m_{i,h}(t) - \beta_T(t_E)] \quad for \;\; h \neq BMU(t-1)$$

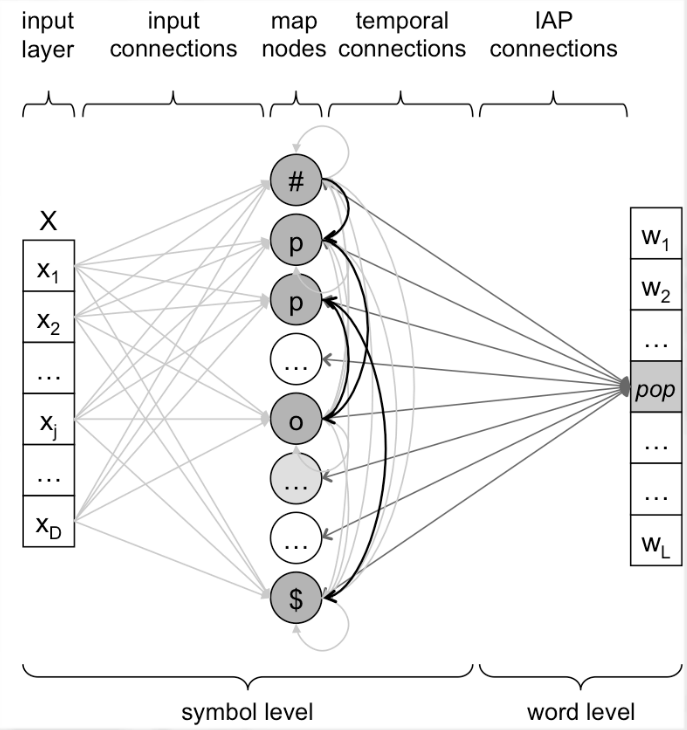


time & frequency: Rescorla & Wagner rules

- for any cue **C** and response ***R***, their association strength
  - grows if C often precedes *R*
    - token freq **entrenchment**
  - decreases if *R* is often preceded by a symbol other than C
    - **competition**; the larger the set of possible cues for *R* the less important they are individually
  - decreases if C is often followed by a response other than *R*
    - **predictivity**: the larger the set of responses to C the weaker its predictivity

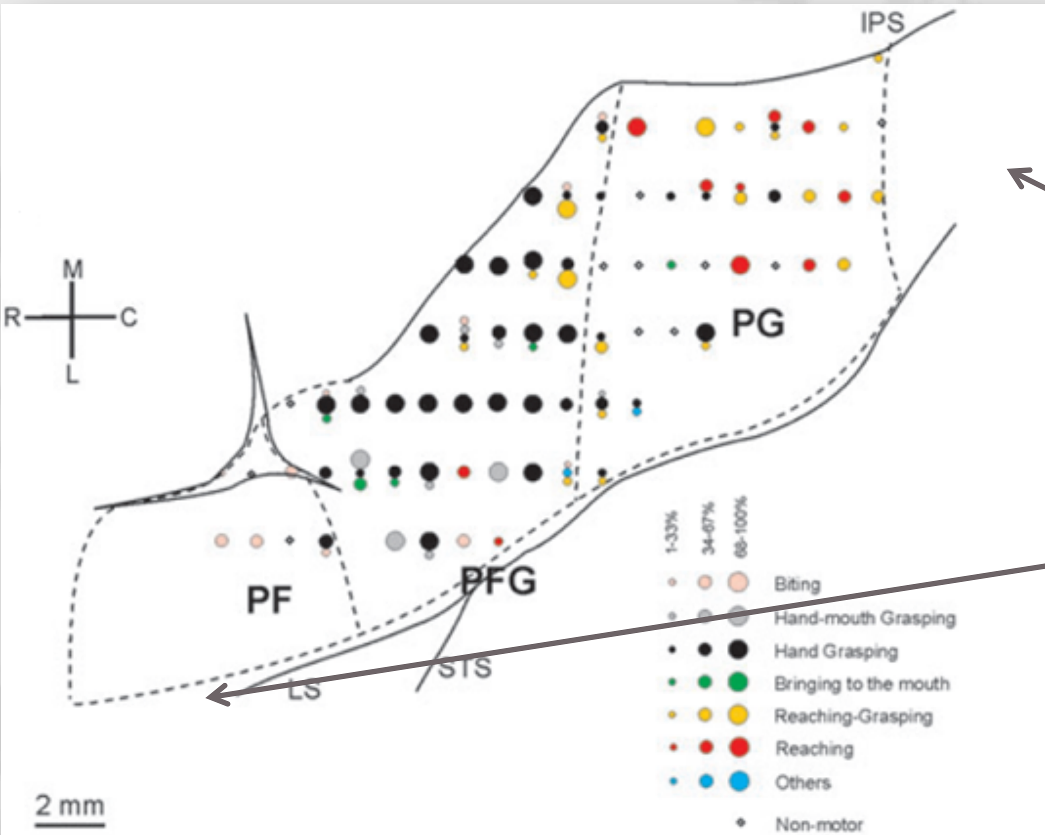
functional principles

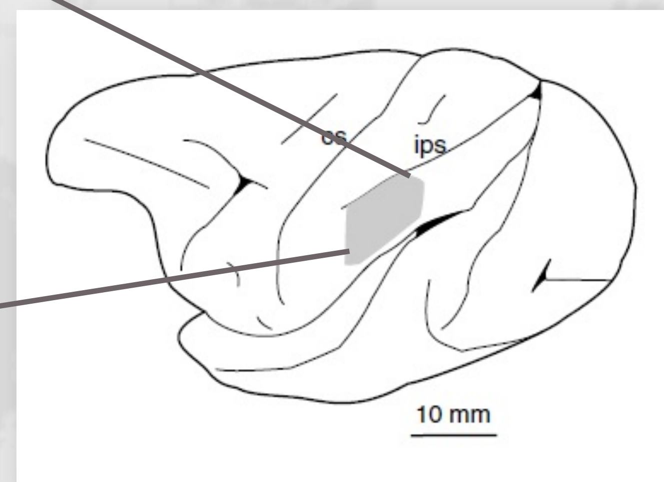
- **competition**
  - nodes are activated **concurrently** but only one wins
- **synchronisation**
  - winning nodes in succession get more and more strongly connected, potentiation being proportional to input frequency
- **specialisation vs. blending**
  - high-frequency and isolated words tend to be processed by **specialised** BMU chains
  - low-frequency input words that are surrounded by many neighbours activate "**blended**" BMU chains, taking part in the processing of more words



emergence of structure in time: gesture coordination in monkeys (and the lexicon)

time-series and motor-coordination

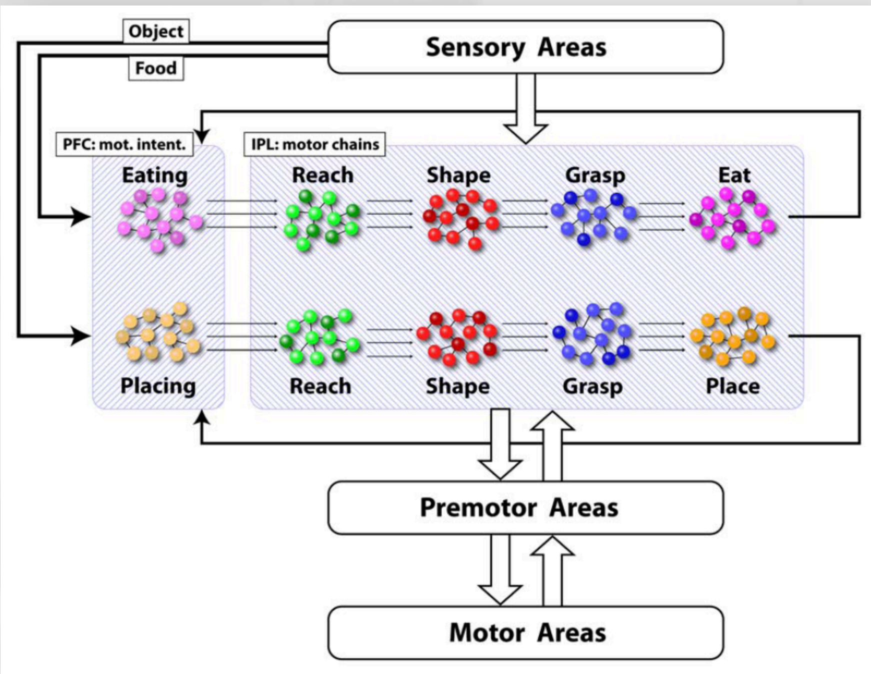




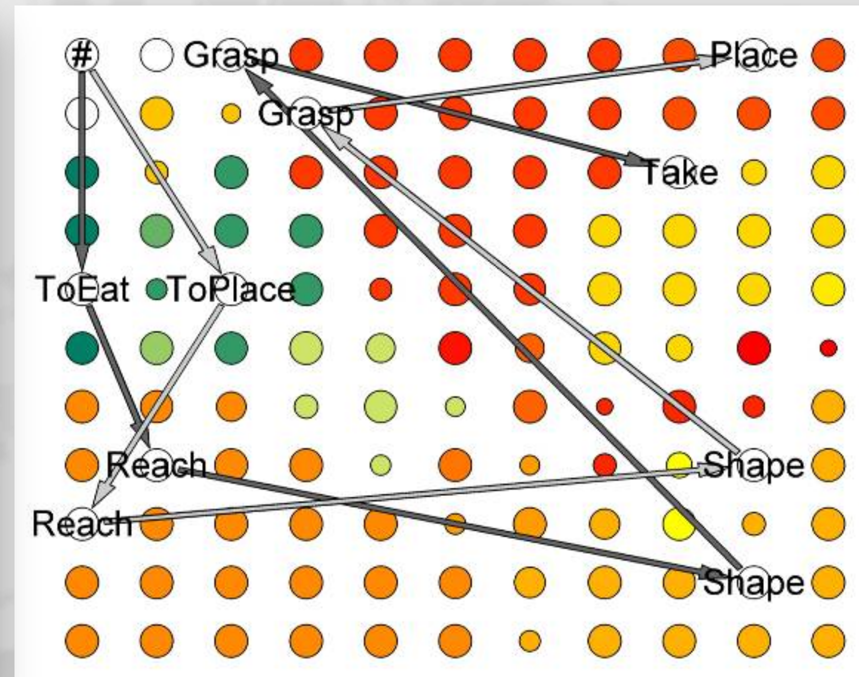
Chersi, F., Ferrari, P.F., Fogassi, L., 2011. Neuronal chains for actions in the parietal lobe: a computational model. PLoS ONE 6

context-dependent gesture coordination

functional organisation

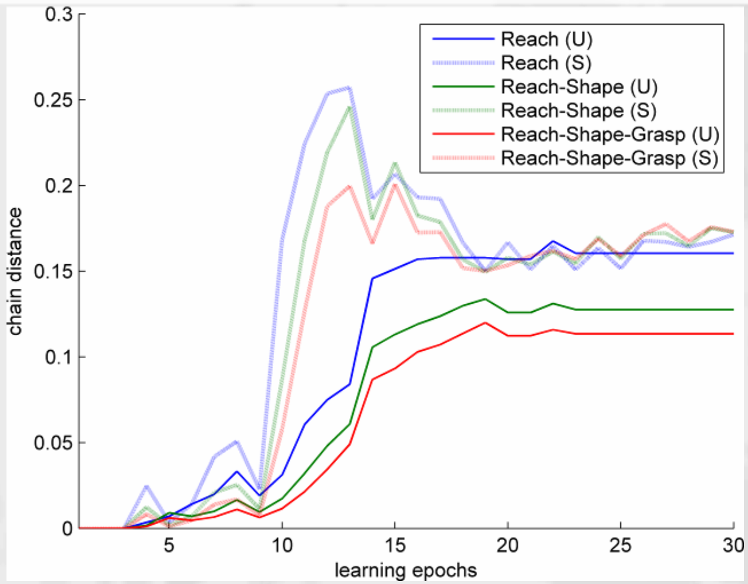


temporal brain map



Chersi, F., Ferro, M., Pezzulo, G., and Pirrelli, V. (2014), "Topological self-organisation and prediction learning support action and lexical chains in the brain", Topics in Cognitive Science (topiCS).

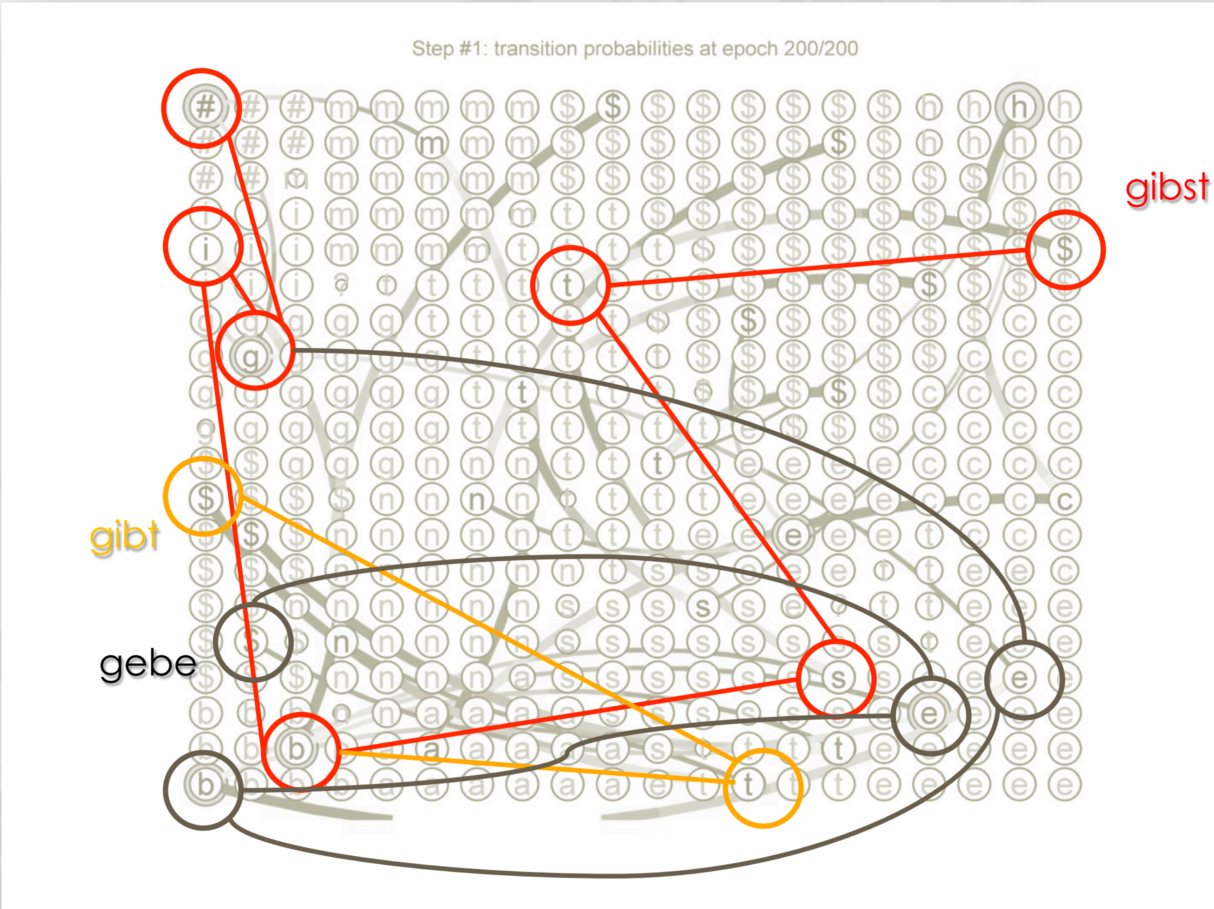
specialisation and sharing



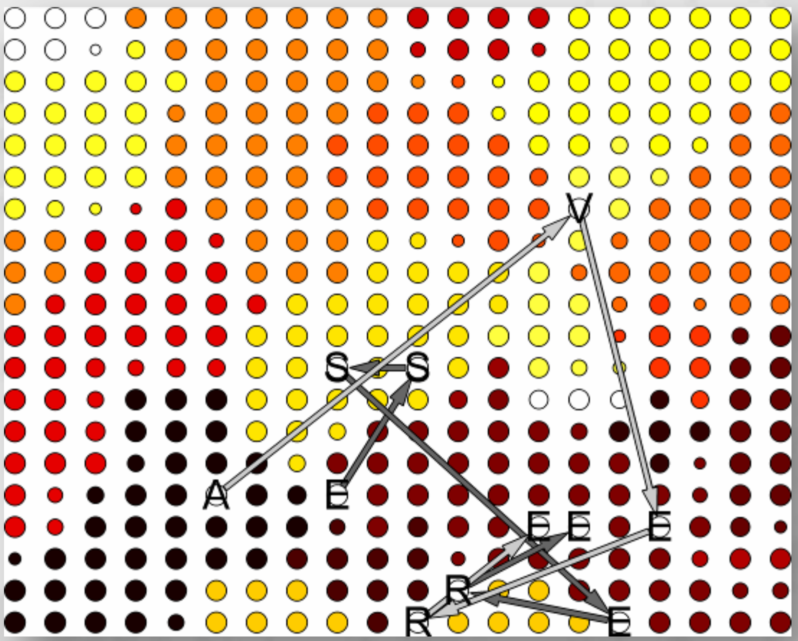
no. chains	freq(To Eat)=55; freq(To Place)=55			freq(To Eat)=100; freq(To Place)=10		
no. nodes	To Eat > To Place:	15.8 (45.3%)	p < 1	To Eat > To Place:	12.6 (88.1%) [72.6%]	p<0.00001
	To Place > To Eat:	19.1 (54.7%)		To Place > To Eat:	1.7 (11.9%) [27.4%]	
	To Eat ≠ To Place:	34.9 (40.1%)		To Eat ≠ To Place:	14.3 (28,6%) [64.2%]	
	To Eat = To Place:	51.1 (59.4%)		To Eat = To Place:	35.7 (71.4%) [35.8%]	
	total:	86.0 (100%)		total:	50.0 (100%)	

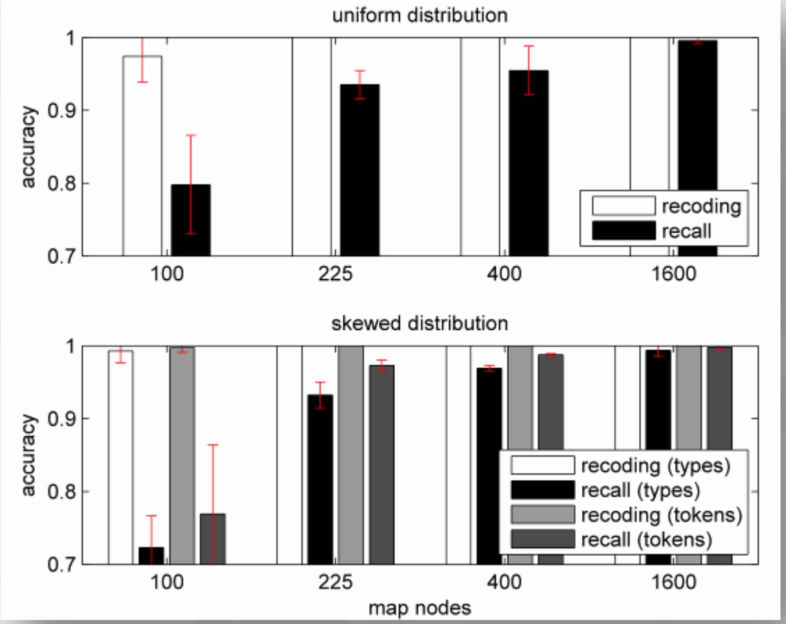
lexical chains: word recoding

#
g
i
b
s
t
$

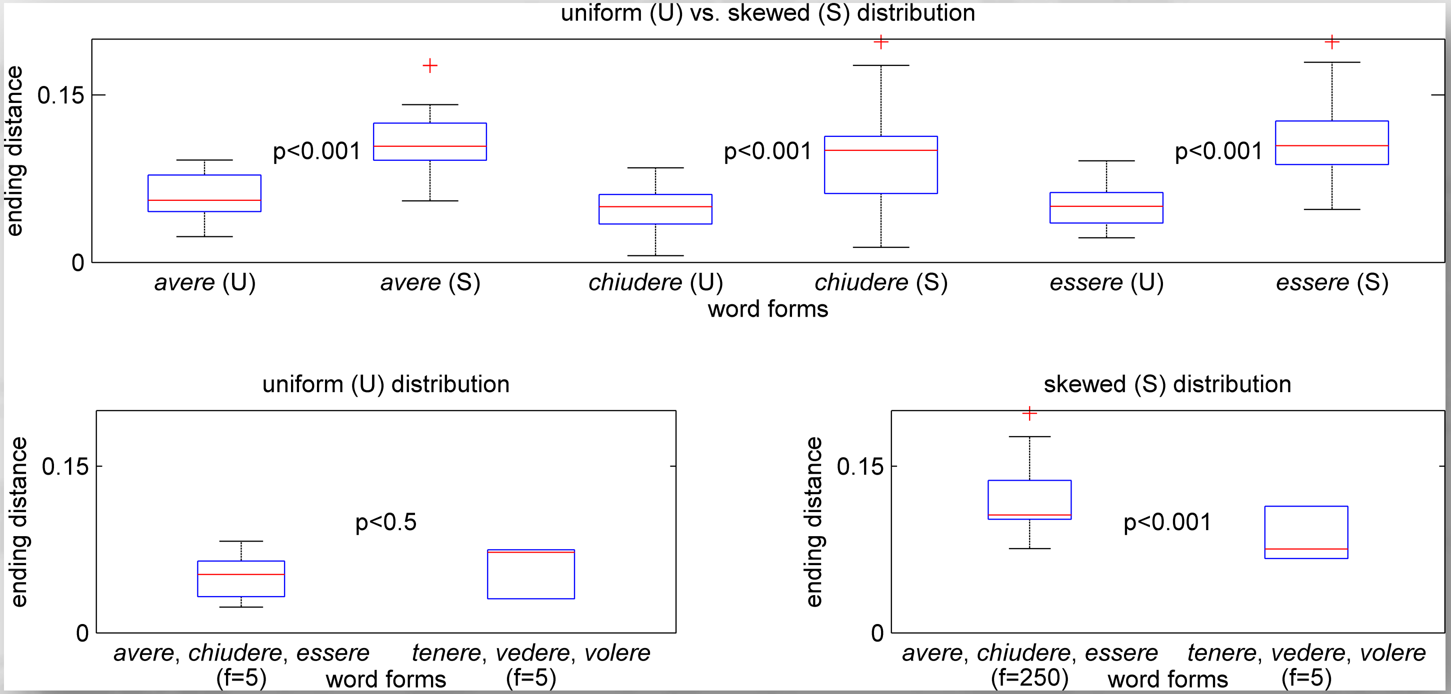


lexical chains in the Italian lexicon





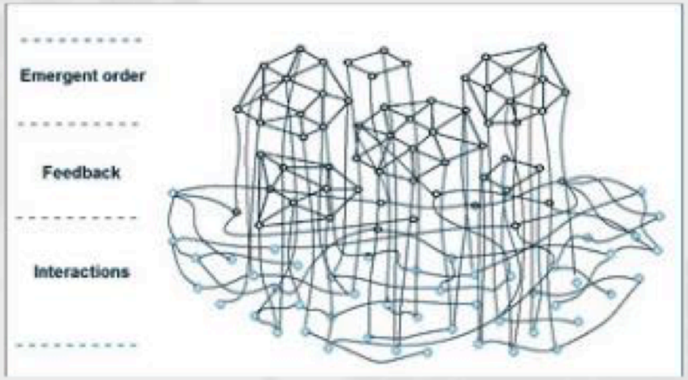
entrenchment



concluding remarks

correlative learning and dynamic memory

- correlative learning can go a long way in developing a notion of **dynamic competitive memory** that blurs the dualism between memory (representations) and processing (rules)



- dynamic memories are gradient, context-sensitive and strongly process-oriented

- at the same time, they enforce a principle of structure-sensitive memory self-organisation through levels of specialised vs. blended connectivity

joint work with…

claudia marzi, marcello ferro

and

fabian chersi, emmanuel keuleers,
petar milin, giovanni pezzulo