A Neuro-Computational Approach to Understanding the Mental Lexicon

Claudia Marzi and Vito Pirrelli

Institute for Computational Linguistics – Italian National Research Council
{claudia.marzi, vito.pirrelli}@ilc.cnr.it

Human lexical knowledge does not appear to be organised to minimise storage, but rather to maximise processing efficiency. The way lexical information is stored reflects the way it is dynamically processed, accessed and retrieved. A detailed analysis of the way words are memorised, of the dynamic interaction between lexical representations and distribution and degrees of regularity in input data, can shed some light on the emergence of structures and relations within fully-stored words. We believe that a bottom-up investigation of low-level memory and processing functions can help understand the cognitive mechanisms that govern word processing in the mental lexicon. Neuro-computational models can play an important role in this inquiry, as they help understand the dynamic nature of lexical representations by establishing an explanatory connection between lexical structures and processing models dictated by the micro-functions of human brain.

Starting from some linguistic, psycholinguistic and neuro-physiological evidence supporting a dynamic view of the mental lexicon as an integrative system, we illustrate Temporal Self Organising-Maps (TSOMs), artificial neural networks that can model such a view by memorising time series of symbolic units (words) as routinized patterns of short-term node activation. On the basis of a simple pool of principles of adaptive Hebbian synchronisation, TSOMs can perceive possible surface relations between word forms and store them by partially overlapping activation patterns, reflecting gradient levels of lexical specificity, from holistic to decompositional lexical representations. We believe that TSOMs offer an algorithmic model of the emergence of high-level, global and language-specific morphological structure through the working of low-


Date submitted: 01/5/15  Date reviewed: 06/25/15
Date confirmed for publication: 08/16/15
©2015 Institute for Cognitive Science, Seoul National University
level, language-aspecific processing functions, thus promising to bridge the persisting gap between high-level principles of grammar architecture (lexicon vs. rules), computational correlates (storage vs. processing) and low-level principles and localisations of brain functions. Extensions of the current TSOM architecture are envisaged and their theoretical implications are discussed.

**Keywords:** Mental lexicon, dynamic storage, parallel distributed processing, hebbian learning, temporal self-organising maps.

1. Introduction

Lexical competence presupposes the fundamental ability to retain sequences of symbolic units in the long-term memory, for them to be accessed and retrieved for language processing more generally. Without assuming the existence of what is commonly referred to as “mental lexicon”, it would be impossible to understand how a child can face the problem of acquiring the large number of lexical “chunks” indispensable for understanding and producing even the simplest sentences. Nonetheless, defining the precise content of the mental lexicon, its basic organisation principles and the cognitive mechanisms serving its use for language processing has proved a challenging scientific endeavour, calling for cross-disciplinary integration of different knowledge domains and methodologies, ranging from theoretical, typological and computational linguistics to cognitive psychology and neuroscience.

For decades, issues of lexical content and organisation have been investigated together with issues of morphological competence, with particular emphasis on aspects of the internal structure of complex words (Bloomfield, 1933; Bloch, 1947; Chomsky & Halle, 1968; Lieber, 1980; Selkirk, 1984). According to the generative view, words are made up out of simpler constituents, or “morphemes”, commonly understood as the irreducible sign-based building blocks of human lexical competence. By decomposing morphologically complex words into their basic constituents, which are recombined into full forms in word production, theoretical morphologists hoped to arrive at the foundations of lexical competence: a redundancy-free repository of sublexical knowledge accounting for the
ways words are mutually related in the speakers’ mind. The theory of speech production developed by Levelt and colleagues (1999) provided a psycholinguistic implementation of this view, where only irreducible forms are stored in the lexicon as separate entries.

However, the existence of a sharp dichotomy between processing and storage has been questioned by a more recent and equally influential tradition in the psycholinguistic literature on morphological competence (e.g. Bybee, 1985; Aitchison, 1987), according to which the mental lexicon cannot be redundancy-free. In fact, the organisation and structure of the morphological lexicon must be maximally functional to lexical access and retrieval, as normal processes of speech production and comprehension must operate within time constraints of the order of hundreds of milliseconds.

This perspective has important consequences on the way we should think about the mental lexicon and its organisational principles. Lexical knowledge, in fact, does not appear to be organised to minimise storage or respond to logical necessity, but it rather tends to maximise processing opportunities (Libben, 2005). Full storage in lexical organisation may respond better to some tasks (e.g. lexical decision), while lexical decomposition can be a more efficient strategy for other tasks (e.g. lexical acquisition). In a biological system like the brain, the best possible solution is to entertain both strategies at the same time, and make them compete on different tasks. This view has been confirmed by empirical findings.

It has been shown, for example, that surface word relations constitute a fundamental domain of morphological competence, with particular emphasis on the interplay between token frequency, family frequency and family size effects within morphologically-based word families (Baayen et al., 1997; Taft, 1979; Hay, 2001; Ford et al., 2003; Lüdeling & De Jong, 2002; Moscoso et al., 2004; Stemberger & Middleton Setchell, 2003; Tabak et al., 2005). Nonetheless, more than just lexical storage must be involved in word processing, as suggested by priming effects between false morphological friends (or pseudo-derivatives) such as broth and brother (Frost et al., 1997; Rastle et al., 2004; Post et al., 2008). As soon as an input sequence is perceived as decomposable into morphological formatives, word decomposition takes place automatically, prior to (or concurrently
with lexical look-up. Word processing is hypothesized to be the outcome of simultaneously activating patterns of lexical knowledge reflecting redundant distributional regularities in input data at the phonological, morphosyntactic and morpho-semantic levels. Different lexical areas devoted to word processing must thus contain redundantly specified knowledge, to maximize the opportunity of using both general and specific information simultaneously (Libben, 2010; Post et al., 2008).

In the last two decades, these lines of argument have clearly developed into a view of the lexicon as a dynamic memory system (e.g. Elman, 1995, 2004; Li, 2009): what is called lexical information is the resulting combination of various dynamic properties, such as the relation of a particular acquired stimulus to other co-occurring stimuli in a certain time window (syntagmatic relation), or the possible relation of the same stimulus to other competing stimuli in complementary distribution (paradigmatic relation). From this perspective, a detailed analysis of the dynamic interaction between lexical representations and distribution, perceptual discriminability and degrees of regularity in input data can shed some light on the emergence of structure in the mental lexicon.

Recent neuro-physiological evidence provides strong support to a highly distributed, multi-level and integrative view of lexical competence. Acquisitions on the neuro-functional architecture for language processing (Catani et al., 2005; Shalom & Poeppel, 2008; Friederici, 2012; among others), unveiling the functional organisation of the perisylvian brain network, support a subdivision of labour into general-purpose low-level operations (short-term storage, long-term storage, sensory-motor mapping, rehearsal, unit integration, unit analysis, executive control, time-series processing). Several brain areas are selectively specialised to support the processing of a large variety of input representations (images, music tones, phonological segments, morphs, words and phrases). Different high-level effects in the operation of this common pool of processes are a by-product of the specific nature of input representations.

In addition, advances in understanding the neuro-anatomical bases of working memory\(^1\) (Wilson, 2001; D’Esposito, 2007; Ma et al., 2014) show

\(^1\) Working Memory (WM) relates to the temporary active retention of information through
that the brain areas involved in both language processing and working memory exhibit considerable overlapping, and that their corresponding low-level functions are mutually implicated. Accordingly, verbal working memory can be seen as the by-product of a dynamic form of sensory-motor integration, with the storage component being located in Wernicke's area (Hickok & Poeppel, 2004; Shalom & Poeppel, 2008). Under this view, working memory consists of the transient activation of long-term memory structures even in the absence of external stimuli, which is controlled and maintained by the integration of auditory-motor circuits in the perisylvian network. An important consequence of this view is that it blurs the distinction between word processing and word memory: structures that are repeatedly activated in processing an input word are the same units responsible for its stored representation. Thus, the deep neuro-physiological interconnection between lexical acquisition, processing, serial cognition and memory for time series shades the distinction between storage and computation to suggest that memorisation and processing are, in fact, two sides of the same coin.

As theoretical scenarios become more complex and causal relations so intricate to defy analytic solutions, computational modelling can be of considerable use to achieve cumulative scientific progress (Norris, 2005), and gain a deeper understanding of how brain areas interact in carrying out a given task, and how information flow translates into behavioural responses.

In line with this perspective, in the present paper we propose a detailed, algorithmic view of morphological competence as a function of the sensory system. It also describes the short-term internal representation of retrieved information that is already stored. Firstly conceptualised by Baddeley (1986) as a brain system providing temporary maintenance and manipulation of the necessary information for complex cognitive tasks such as language comprehension and acquisition, WM is currently viewed as a network that emerges from the functional, goal-directed interaction between different brain regions (including the prefontal cortex). It is generally considered to have limited capacity (or “span”), depending on the category of information (Cowan 2001). Its span is understood to depend on domain-specific chunking and rehearsal mechanisms facilitating storage and a domain-general capability allowing for cognitive control and executive attention.
integration of low-level processes, whereby morphological representations are determined by probabilistically recurrent processing strategies that are routinized and stored in long-term memory.

There is growing awareness that computer models can contribute considerably to a precise characterisation and assessment of functional models of the mental lexicon, since they provide experimental ways of testing the behaviour of a complex system as a function of different settings of configuration parameters and input conditions (Pirrelli, 2007). Furthermore, many real systems can be modelled as artificial neural networks, where the elements of the system are nodes, and interactions between elements can be modelled as adaptive processes giving rise to emergent behaviour. Networks thus represent the support of many complex systems, and their theoretical and computational analysis makes it possible to gain insights into their organisational and functional dynamics.

In this perspective, artificial neural networks, and self-organising connectionist models in particular, can do something more than just hypothesis testing. They promise to provide the intermediate level of scientific inquiry that bridges the gap between low-level, interactive brain processes and high-level language knowledge and language behaviour. By developing explicit, biologically-inspired models, computer simulations can help understand more about the emergence of complex lexical representations from highly-interconnected relations taking place at word level across different time scales. Here, we illustrate some of the virtues of this approach, together with its current limitations and future perspectives, to show that linguistic requirements on lexical representations, psychocognitive behavioural evidence and neuro-functional constraints can be integrated within a unitary explanatory model. Theoretical linguists have investigated what sort of linguistic representations and features speakers must acquire for mastering language, and psycholinguists and neuro-linguists have provided evidence of when and where representations and features are evoked and used in the brain. Neuro-computational models can tell us how this can come about algorithmically, by implementing detailed models of the interaction between linguistic knowledge and psychocognitive and neuro-functional constraints. To make our discussion more concrete, we will focus on a specific neuro-computational model for word
processing and storage known as Temporal Self-Organising Map (or TSOM for short, Ferro et al., 2011; Marzi et al., 2012; Marzi et al., 2014; Pirrelli et al., 2014), which exhibits interesting properties of dynamic self-organisation of morphological knowledge based on correlative learning and competitive activation. In particular, we will deal with issues of lexical competition and organisation within so-called inflectional paradigms, i.e. families of inflected variants of the same lexical exponent (e.g. walk, walks, walked, walking), and inflectional classes, i.e. families of similarly inflected forms (e.g. walked, loved, typed, or stung, swung, wrung). We intend to show that interconnectivity between concurrently fully memorised forms can go a long way in accounting for effects of sensitivity to graded morphological structure, frequency by regularity interaction and family-based facilitation/inhibition.

It should be appreciated that the role of TSOMs in the present paper is to help illustrate the advantages and prospective requirements of the methodological approach advocated here, rather than hold them up as the best possible model compared with other models with similar dynamic properties (e.g. Li, Zhao & MacWhinney, 2007; Mayor & Plunkett, 2010; Althaus & Mareschal, 2013; among others). Actually, by narrowing down our investigation and focusing on a particular computer model only, we hope to be able to present our proposal and methodological points in a clear and convincing way. Most of the points argued here will likely hold, with some qualifications, for other existing models.

In what follows, we first provide an overview of recent psycho-cognitive and neuro-linguistic evidence supporting an integrative, dynamic model of the mental lexicon as the result of the functional integration of low-level language-aspecific processing functions and lexical, language-specific properties emerging from concurrent memorisation of full forms (section 2). We then move on to considering the requirements of a neuro-computational model that can support such a view (section 3), and discuss experimental results obtained by running a specific computational architecture based on Temporal Self-Organising maps (section 4). Finally, (in section 5) we elaborate some methodological points and lay the foundations of a more comprehensive architecture.
2. Psycho-Cognitive Evidence and Neuro-Linguistic Correlates

Any attempt at modelling the mental lexicon must take issues of storage seriously. Word storage plays a fundamental role in both word acquisition and processing, and the way words are structured in our long-term memory is key to understanding the mechanisms governing word access. If we abstract away from differences in modality-specific input stimuli (e.g. orthographic vs. phonological), a common core of computational principles of correlative learning appears to determine the way words are coded as orderly series and co-activate and compete in processing. Since words are permanently coded in our long-term memory as neuron activation patterns that sequentially fire, they can be conceptualised as time series of symbols (be they letters or sounds), whose receptors are time-bound to one another through associative connections. For sure, the way word-based activation patterns competitively interact may vary depending on the specific nature of the input stimulus (e.g. acoustic vs. visual) and on the task (e.g. word production vs. word recognition). Nonetheless, investigating the functional principles underpinning dynamic storage of words has proved an extremely productive line of scientific inquiry, whose main acquisitions are summarised below.

2.1. Parallel Distributed Processing

Most psycholinguistic models of the mental lexicon are based on the fundamental hypothesis – confirmed by neuro–functional evidence² –

---
² Functional magnetic resonance imaging (fMRI) studies on speech comprehension report activation in four key regions: anterior and posterior parts of the left middle temporal gyrus, bilateral anterior temporal poles, left angular gyrus, and the posterior cingulate/precuneus. Semantic processing of familiar auditory stimuli activates a distributed set of regions that surround the ventral, anterior, and posterior borders of the perisylvian regions supporting pre-lexical auditory speech processing. The extended distribution of semantic activations suggests that there are numerous pathways supporting speech perception and comprehension. It has been suggested (Price, 2000) that acoustic analysis of words is conducted in the superior temporal cortex, visual analysis of written words in the posterior temporal cortex and temporo-occipital cortex, and semantic representations in a network
that the lexical processor consists of a network of parallel processing units (functionally equivalent to neuron clusters) selectively firing in response to sensory stimuli (McClelland & Elman, 1986; Norris, 1994; Luce & Pisoni, 1998). In processing the input stream, sensory information initiates concurrent activation of the appropriate nodes that respond to features/units of the input as they unfold through time. When activation spreads to the lexical level, multiple lexical candidates are co-activated and compete with each other for final selection. Goodness-of-fit criteria guide the activation towards the optimal candidate, which is eventually singled out as the final winner. These basic assumptions appear to capture aspects of the dynamicity of mental processes underpinning lexical competence (e.g. Bybee, 1995) and are supported by large-scale frameworks of the functional anatomy of language that integrate neuro-psychological, neuroimaging and psycho-linguistic data (e.g. Hickok & Poeppel, 2004; Shalom & Poeppel, 2008).

Neuroimaging evidence suggests that the neural circuits supporting language functions must connect the language-relevant brain regions located in the inferior frontal cortex, including Broca’s area, and the superior temporal cortex, including Wernicke’s area. Different brain regions exhibit persistent neural activity during active maintenance of task-relevant information, and are recruited in a unique network depending on the type of information being actively maintained (D’Esposito, 2007). The location of pre-lexical activation appears to be determined by the task demands. For example, when an articulatory response is required, top-down expectations from motor programs may guide pre-lexical processing in a dorsal processing direction. When a lexical decision is required, top-down expectations from semantic knowledge may affect pre-lexical activation in a ventral direction. In this way, the cortical networks supporting language comprehension are dynamically determined by the task and context (Saur et al., 2008).

At the same time, independent neuroimaging findings have highlighted the role of prefrontal regions in the sustained maintenance and manipulation of information over short delays, or working memory (WM), and in the including the angular gyrus and the anterior inferior temporal cortex.
formation and retrieval of memories for events, or episodic long-term memory (LTM).\textsuperscript{3} This overall picture was confirmed by Diffusion Tensor Magnetic Resonance imaging data of the left hemisphere (Catani et al., 2005), providing neuro-anatomical evidence of a bidirectional perisylvian pathway from the Superior Temporal Gyrus (or Wernicke’s area) to Broca’s area through the Inferior Parietal Lobe. The bidirectional pathway defines the neuro-cognitive substrate to the retention of sequences of linguistic units and orosensory goals for their covert vocalisation in working memory. In keeping with Baddeley’s phonological loop hypothesis, motor-to-auditory mappings are used to activate auditory representations of speech through controlled articulatory rehearsal in frontal circuits (Gathercole & Baddeley, 1989; Papagno et al., 1991).

2.2. Dynamic Storage

The evidence reported so far establishes a close connection between lexical processing and working memory in terms of the anatomical arrangement of the neural networks involved. Linguistic units that are frequently sequenced together are consolidated (entrenched) in Wernicke’s area, where they are accessed and executed as highly automatized routines. This increases fluency and articulation rate and accounts for the memory effect known as “chunking”: the presence of one or more routinized chunks in a sequence of symbols to be processed augments the capacity of our working memory, where more fluent chunks can be rehearsed more quickly (Miller, 1956; Baddeley, 1986; Cowan, 2001). Chunking effects, together with recent psychophysical studies showing that recall precision declines continuously as the number of items to be remembered increases (Ma et al., 2014), led scholars to revise Baddeley’s original hypothesis that working-memory consists of a short-term dedicated buffer area, made up out of a fixed number of memory slots.

\textsuperscript{3} Results show that the same bilateral ventro-lateral prefrontal regions (at or near BA 6, 44, 45, and 47) and dorsolateral prefrontal regions (BA 9, 46) are involved in both encoding and recognition within the context of WM and LTM tasks (D’Esposito et al., 1999; Wagner, 1999; among others).
Accordingly, working memory is no longer conceptualised as a unitary dedicated system, but rather, following independent evidence and suggestions from several researchers (e.g. D'Esposito, 2007; Hickok et al., 2003; Wilson, 2001; among others), as a dynamic form of sensory-motor integration, with different circuits being involved and interconnected depending on the goal-directed task and the information to be processed. Active maintenance involves the recruitment of the same circuitry that represents the stored information itself, with different circuits for different types of somato-sensory information. It has also been suggested (Catani et al., 2005) that highly automatized routines may activate a more direct path mapping acoustic time-series onto articulatory sequences when deeply entrenched sound sequences are to be vocalised. Conversely, novel sequences are more difficult to repeat and they may take indirect routes, involving motor programs, analysis of underlying structure and conscious effort, to be properly articulated.

But fluency and articulation rate can explain only part of the evidence. Long-term memory effects appear to interact with short-term memory efficiency. A sequence of words concatenated in a meaningful sentence is shown to be retained much more easily than a sequence of unrelated words, independently of their length (Papagno et al., 1991; Hulme et al., 1997; Vitevich, 2002; among others). The effect indicates that probabilistically more supported sequences are recalled more easily than unpredictable ones. Likewise, an input sequence containing many familiar chunks is more likely to be actively retained in working memory even if it exceeds the short-term span (Hulme et al., 1991; Gathercole et al., 1991). Thus, chunk integration augments the capacity of the short-term memory system (working memory) to maintain and manipulate complex input sequences. By recursive application of chunking, once a temporal sequence of items is perceived as a single unit, it may be part of complex sequences of chunks, thereby producing levels of hierarchical organisation of the input stream: what is perceived as a temporal sequence of items at one level may be perceived as a single unit on a higher level, to become part of more complex sequences (Hay & Baayen, 2002). Furthermore, as sequences of chunks are not equally likely, with some inter-chunk transitions being more frequent and entrenched than others, perception of chunk-based structure is
inherently graded (Hay & Baayen, 2005).

The correlation between frequency of input and perception of internal structure of word forms suggests that access representations in the lexicon depend on how inputs are differentially processed in serial perception and storage. The way information is stored reflects the way it is dynamically represented, accessed and retrieved.

In this perspective, we believe that the way words are memorised can help understand the cognitive mechanisms that govern word processing and the organisation of words in the mental lexicon. In particular, a bottom-up investigation of low-level memory and processing functions (e.g. activation, recoding, entrenchment, competition, access, reinstatement and recall) should assess and explain their involvement in language processing.

In what follows, we will consider in some detail how the evidence reviewed so far impacts on requirements for neuro-computational models of the mental lexicon. How are word stimuli internally represented in the brain? How do different representation and processing levels interact with each other in a neuro-biologically inspired architecture? How much of lexical organisation is amenable to general principles of serial memory and cognition and how much should rather be attributed to the operation of specific morphological principles that vary typologically across the world languages?

3. Neuro-Computational Modelling of the Mental Lexicon

Computational models of the mental lexicon are not surrogates for theories of lexical representation, but ways to put theories to the challenging test of a fully spelled-out algorithmic implementation. In doing so, computational models must make explicit somewhat arbitrary decisions that theories often take for granted. Unlike theories, computational models cannot be given the benefit of doubt. Anything that is assumed as a precondition for an implemented model to function must be acknowledged and be given reason for. Every empirical aspect of data coding and testing must be made explicit and justified. Every hypothesised representational structure must be proved compatible with input conditions, processing constraints and task-oriented requirements. Finally, both qualitative and quantitative predictions made by
the model must be pitted against evidence of human behaviour and brain physiology.

All these requirements are critical when we start from the assumption, inescapable in dealing with lexical acquisition, that words are not externally given objects, whose perceptual structure is “out there” for speakers to acquire, but internalised representations, whose hidden nature is a function of perception and processing strategies.

This is true, in an even deeper sense, when it comes to neurocomputational models of the lexicon, i.e. to computer simulations of processes of lexical organisation based on, and concerned with neurophysiologically-grounded principles of information processing. Given the persisting gap between high-level principles of grammar architecture (lexicon vs. rules), their computational correlates (storage vs. processing) and low-level principles and localisations of brain functions, neurocomputational modelling must provide those intermediate-level tools and protocols establishing the relation of language to cognitive neuroscience.

Neuro-computational models can help us understand the very nature of internalised lexical representations by establishing an explanatory connection between assumed lexical structures, as argued for by linguists, and processing models dictated by the micro-functions of human brain. Ultimately, no linguistic structure should be assumed to exist unless the model can offer an explanation of how it comes about and is implemented in the human brain in the first place. This causal connection can in principle be investigated, since neuro-computational simulations can investigate what is not amenable to symbolic manipulation and analytic explanation, e.g. the non-deterministic behaviour of a dynamic system whose many conditioning factors simultaneously change and mutually interact with time.

Likewise, neuro-computational models are expected to reproduce aspects of the interaction between long-term memory and working memory, such as effects of long-term memory chunks on working memory capacity, or long-term entrenchment of input stimuli through reverberatory mechanisms in working memory (phonological loop), lending support to interactive views of lexical memories. Similarly, evidence that active maintenance or storage of task-relevant information involves the very same brain circuitry that sustains the perceptual representation of information (D'Esposito,
Claudia Marzi and Vito Pirrelli (2007) supports a neuro-computational architecture where working memory information defines the subset of long-term information within attentional focus during processing. An input word causes an activation state over some cortical areas of the brain, where other non-target words happen to be co-activated. This prompts (i) a possibly noisy representation of the target word, and (ii) competition among concurrent but distinct activation patterns associatively related to the target word. In this formulation, it is not the number of remembered items, but rather the quality and precision of the levels of activation of relevant representations that define the limits of working memory capacity. Performance on working memory tasks is crucially determined by the discriminability of activation levels between relevant and irrelevant representations (Kimberg et al., 1997). This accords well with general models of perception, suggesting a neuro-computational approach whereby “[…] rather than putting word knowledge into a passive storage […], words might be thought of in the same way that one thinks of other kinds of sensory stimuli” (Elman, 2004: 301).

3.1. Temporal Self-Organising Maps

![Diagram of a TSOM](image)

**Figure 1.** Outline architecture of a TSOM. Map nodes are fully connected. For simplicity, only connections from and to the map central node are shown.

In the neuro-computational framework of TSOMs (Ferro et al., 2010; Ferro
et al., 2011; Marzi et al., 2012; Marzi et al., 2014; Marzi et al., 2015; Pirrelli et al., 2014), word processing and lexical acquisition are implemented as recoding and storage strategies for time-series of symbolic units, dependent on both language-specific factors and extra-linguistic cognitive functions such as lexical organisation, lexical access and recall, input-output representations, and memory self-organisation.

In its simplest instantiation (Figure 1), a TSOM consists of a grid of memory nodes, each of which is acting as a receptor with a Gaussian spreading activation function. Nodes present two levels of synaptic connectivity. Input connections, or “what” connections, convey to map nodes an input signal encoded on the vector input layer. Hebbian connections, or “when” connections, mutually link all map nodes. The weights on Hebbian connections reflect a probabilistic expectation that a post-synaptic node is activated immediately after the pre-synaptic node is activated.

The processing response of a TSOM to an input signal at time $t$ is a pattern of activation of all map nodes, whose maximally activated node, or winning node, is known as Best Matching Unit (BMU). Activation of each node in the pattern is a nonlinear function of the input signal, weighted by “what” connections, and the activation pattern at time $t-1$, weighted by “when” connections, in line with analogous mechanisms observed in cortical areas involved in the classification of sensory data (Aflalo & Graziano, 2006). Thus, any activation pattern responding to a specific input signal contains, besides the BMU for the current signal, a number of other concurrently activated nodes, corresponding to the BMUs for either similar or highly expected input signals. The stronger the level of co-activation, the more difficult it is for the current BMU to overcome the competition of co-activated nodes.

---

4 Hebbian principles define a basic mechanism whereby connections between neurons change through adaptation. Inspired in large part by the dynamics of biological systems, the algorithm representing these changes is based on Hebb’s postulate (1949), which states that when a pre-synaptic neuron fires, if it frequently leads to the firing of the post-synaptic neuron, their synaptic junction is strengthened. This will strengthen the connection between two sequential neurons/nodes, while less-effective synaptic connections are weakened.
An input word is administered to a TSOM as a sequence of units (e.g. sounds, letters, syllables, letter chunks, etc., depending on the level of representation and the task), each presented on the input layer one at a time tick. Start and end of an input word are marked by ‘#’ and ‘$’ respectively. A word like pop is thus encoded on the input layer by first presenting the vector code representation for ‘#’, followed by ‘p’, followed by ‘o’, followed by another input vector for ‘p’ (identical to the previous instantiation of the same symbol), finally followed by the vector for “$”. Upon being presented with another input vector for ‘#’, the map resets its pattern of activation, thus losing memory of the last presented word. The map’s response to an input word thus corresponds to the vector sum of the activation patterns triggered by each symbol in the word, or Integrated Activation Pattern (IAP). An example is illustrated in Figure 2 for the string #pop$, where arcs show Hebbian connections between consecutively activated BMUs. Note that two different BMUs are labelled with p, one for each input p, showing that the map developed two different time-bound responses to the same input symbol.

Figure 2. An Integrated Activation Pattern (IAP) associated with the input string #pop$. Directed arcs stand for forward Hebbian connections between BMUs. Shades of grey depict levels of node activation.
For our present purposes, training a TSOM consists in showing a random sequence of words, each sampled according to its frequency distribution. During training, weights on both connectivity layers are adjusted in an experience-dependent way, based on the distribution of time-bound signals, for nodes to become increasingly sensitive to specific input signals in specific temporal contexts. What a TSOM does is to cache its best responding activation patterns by strengthening “what” and “when” connections to winning nodes and weakening “what” and “when” connections to losing node. Due to training, after an initial period of random variability in the map’s response, a TSOM increasingly stabilises its signal-response mapping, until a learning stage is reached where a certain input signal in context is repeatedly and systematically associated with a stable activation pattern. Like a child acquiring a lexicon, the map gradually develops increasingly more detailed and specific IAPs for words in the training set. The incremental process starts with probabilistically more supported (see Figure 3) and shorter word chunks, to acquire longer and less supported sequences only at later stages.

Figure 3. Mean frequency of correctly recalled German verb forms over learning epochs: more frequent word tokens tend to be acquired earlier. (Adapted from Marzi et al., 2014)

Since the IAP of an input word contains information of all symbols making up that word and their context, the map can reinstate, from the IAP, the sequence of units that generated it, starting from the ‘#’ node as the first BMU, and letting activation propagate through the path of Hebbian
connections from ‘#’ to ‘$’. At each time tick, the current BMU is calculated and its associated symbol is output. We can thus say that in TSOMs lexical storage is based on recurrent activation patterns that get integrated on the map as routinized circuits. Ideally, each such circuit should contain information on how the original input sequence can be reinstated as a time series of units.

4. TSOMs at Work

TSOMs do not generate an output form from an input form (e.g. walking from its base walk) as customary with traditional multi-layered perceptrons (e.g. Rumelhart & McClelland, 1986). A defining feature of TSOMs is the ability to store input forms by recoding them into maximally efficient activation patterns. Hence TSOMs are, primarily, models of lexical memory in a strict sense. But there is more to it than just storage.

In TSOMs, individual input word forms are assumed to simply consist of a linear arrangement of time-bound units, vector-coded on the input layer at a discrete time pace. In making this choice, we take an abstract, minimalist view on matters of input representation. Differences in memory self-organisation and internally stored representations are the by-product of acquired sensitivity of map nodes to recurrent (sub-)lexical chunks. This two-level approach is based on the intuition that peripheral representations on the input layer are mapped onto less peripheral representations reflecting how the map internalises and perceives recurrent time-bound series. Lexical structure is not vector-coded on the input layer. It is only at the map level that it is discovered, due to repetition of frequently shared sequences in the input (Bybee, 2002). If two input word forms like macht and gemacht share a redundant chunk, they will tend to recruit and activate an overlapping pool of nodes. We will return to the role of frequency, sequentiality and structure in lexical organisation in the following sections.

In our view, this represents a biologically plausible solution to the word-coding problem, and avoids the notorious principled difficulties of conjunctive representations (e.g. Wickelcoding) in multi-layered perceptrons, which presuppose considerable built-in knowledge of the morphology to be acquired (Sibley et al. 2008).
From this perspective, based on a simple architecture and a sparse set of principles of Hebbian synchronisation, TSOMs can simulate global processes of lexical organisation and emergence of lexical structure through interaction of low-level, local processing functions such as (spreading) activation, recoding, entrenchment, competition, access, reinstatement and recall. They can thus provide an intermediate-level computational framework integrating general-purpose, language-aspecific functions for serial processing with high-level, language-specific effects of lexical organisation. In what follows, we are going to overview some of this evidence, and frame it in the context of more general cognitive issues.

4.1. Storage and Processing

TSOMs exhibit a short-term and a long-term dynamic. The short-term dynamic corresponds to the processing response of a map being presented with an input stimulus at a given point in time, and consists in a pattern of node activation quickly decaying over time. When the input is a sequence of stimuli stretching over more time ticks, the short-term dynamic unfolds throughout the duration of the input sequence, with early stimuli getting integrated with upcoming stimuli in the same IAP. Temporary activation of an IAP is understood to last a few more instants after the end of stimulus presentation, to simulate a working memory buffer where later stimuli reach their activation peak when earlier stimuli are already decaying (as claimed by models of serial order based on “competitive queuing”, e.g. Houghton, 1990).

This short-term processing dynamic is accompanied by a long-term dynamic that drives memory and learning. At each time tick the pattern of nodes responding to the current input gets consolidated by adjusting both “what” and “when” connection weights. Connections to the current BMU are strengthened, and connections to non-BMU nodes are weakened.

As a result, the next time the same stimulus is input again in the same context, the BMU will respond to it more strongly. This prompts a long-term process of selective specialisation, whereby a repeatedly activated pattern is stored and systematically associated with an input signal in context. At the same time, weight adjustment spreads radially with a Gaussian function
centred on the current BMU to involve other neighbouring nodes. Parallel spreading activation makes room for information sharing and training dependence between topologically neighbouring nodes, which respond alike to identical input stimuli embedded in different contexts. In the end, some BMUs may be specialised for specific IAPs only, while some others may respond to a broader range of input conditions that blend the requirements of several activation patterns. This is modulated by frequency distribution (see sections 4.2 e 4.3) and amount of formal redundancy in the training data (see section 4.4).

It should be emphasised that working memory is not implemented here as a buffer physically distinct from long-term memory structures. Rather, the role of working memory is to temporarily sustain and manage activation patterns involving the same pools of nodes that support long-term information.

During the learning process, the relationship between short-term activation and long-term adjustment establishes a two-fold dynamic between processing and storage. On the one hand, processing depends on storage, as it consists in the temporary (short-term) re-activation of a pattern of memory nodes keeping long-term information on their connections. On the other hand, long-term storage depends on processing and it consists in caching successful, routinized processing responses (short-term activation patterns).

The two-fold dynamic is illustrated in Figure 4, depicting how well IAPs work for processing and storage when a TSOM is learning 750 German verb forms (15 forms for each of the 50 most frequent paradigms), whose distribution is a function of word frequencies in Celex (Baayen et al., 1995). In the plot, processing accuracy is calculated as the percentage of training word types that are recoded accurately by the map. Storage accuracy is measured as the percentage of word types that are recalled accurately from their IAPs. Storage accuracy presupposes processing accuracy, as a word cannot be recalled accurately from its own IAP if its symbols are not recoded accurately in the first place. However, it may be the case that an IAP fails to contain the temporal information needed to reinstate, in the appropriate order, symbols that are faithfully recoded. This is because IAPs require considerable fine-tuning of Hebbian connections before they can
be used as reliable memory traces for words. This explains the time lag
between the two curves and the steeper growing slope in recoding accuracy.

![Figure 4. Recoding and recall accuracy for German verb form types over first 25 learning](image)

### 4.2. Holistic vs. Decompositional Storage

In TSOMs, the amount of memory resources that are devoted to an individual input word is a function of a number of factors. Some factors are related to the input word itself, such as its frequency and typicality (i.e. the extent to which the word is perceived as similar to other words in the lexicon). Some others have to do with the global internal state of the map: e.g. the map’s propensity for spreading information through the network of associative connections, and its epoch-sensitive plasticity (i.e. readiness to adaptively adjust connections).

In particular, frequency plays a fundamental role in mediating the impact of selective specialisation and spreading activation on the map self-organisation. Figure 5 sketches this role. The graph in (a) represents the patterning of connections between nodes at the beginning of epoch 1, when...
all connection weights are uniformly set to 0.5. At the outset, all possible paths through nodes are equally likely to be traversed, since the map is a *tabula rasa*.

The other two graphs represent possible, alternative outcomes of two different training regimes: (b) with high-frequency unbalanced distributions of training data and (c) with low-frequency more balanced distributions of the same data. In (b), strings #ABC$, #BC$ and #B$ are structured in a tree. They are assigned independent representational resources (no overlapping nodes), with no compression of redundant information, while all untrained paths are given a null probability. This corresponds to a situation where all training words are memorised as wholes, with no sharing of (sub-)lexical chunks. Prompted with the unknown string #AB$ (i.e. a string that was never shown in training), the map will not be able to find a connection between A and B, thus failing to re-use stored activation patterns for a new purpose. Unlike (b), the graph in (c) shows several converging paths, with multiple paths varying activated by the same input string. Many possible paths that are only partially activated by training data are given a non-null weight. Thus, the map is ready to recognise or produce strings that never appeared in its training set. For example, #AB$ will first activate an existing #-A connection, to then go through a weak (dashed) A-B connection and an attested B-$ one. New sequences can exploit old paths, and node sharing favours local generalisations.

Graphs (b) and (c) in Figure 5 are just two opposite ends in a continuous range of possible graded outcomes. A map may eventually settle on patterns of connection weights that are intermediate between (b) and (c), as a function of frequency distributions and redundant structures in training data. An interesting implication of this behaviour is that holistic and decompositional storage are not alternative strategies, but rather graded outcomes of the same adaptive learning strategy, whose response is a function of the probabilistic support received by the map from a realistic distribution of symbolic time-series, and of the amount of formal redundancy shared by them.

As we will see in more detail in the following sections, this behaviour illustrates a distributed and graded view of morphological structure as an emergent property of lexical organisation, based on the assumption that...
all words are, in a way, memorised in the lexicon, but not all words are
memorised equally. Likewise, by blurring the distinction between memory
and processing, TSOMs do not require indirect access mechanisms and
intermediate sub-lexical representations for lexical access (e.g. Pinker
& Ullman, 2002; Ullman, 2004), but store holistic and decomposed
representations on the same level of connectivity, providing for an
integrative memory-based account of word processing, access and recall.

![Figure 5](image)

**Figure 5.** Three connectivity patterns, as the outcome of different training
conditions: (a) before training (learning epoch 1); (b) after high-frequency
unbalanced training; (c) after balanced training. Solid black arrows represent
entrenched forward connections (with thicker lines depicting stronger
connections); dashed arrows represent forward connections in partially activated
patterns; faint arcs stand for untrained connections. The hash-marked node (‘#’)
prompts the start of an activation pattern.

### 4.3. Two Dimensions of Lexical Organisation

Selective specialisation defines the bias of a TSOM to develop entrenched
node chains for frequent input sequences, mimicking human sensitivity to
the most typical chunks (be they phonotactic or orthotactic) in a language.
This serial (or *syntagmatic*) dimension of lexical organisation interacts with
another parallel (or *paradigmatic*) dimension of word perception, based on
the observation that when speakers process an input word other non-target
neighbouring words get activated in parallel and compete for selection (Marslen-Wilson, 1987; Norris, 1994; Luce & Pisoni, 1998; among others).

In the psycholinguistic literature on word association, the two dimensions of lexical organisation have traditionally been interpreted in terms of two quantitative correlates: n-gram probability density (i.e. the likelihood that a word form results from concatenation of sublexical chunks of n length) for syntagmatic relations, and lexical density (the number of word forms in the lexicon that are similar to a specific target word) for paradigmatic relations (Bailey & Hahn, 2001). In the framework of network morphology, Bybee (1995) suggests resorting to two distinct computational mechanisms, namely lexical entrenchment and lexical association, independently operating and inversely correlated in the lexicon, to account for these effects. We argue that co-activation/competition based models of lexical acquisition like TSOMs can account for the two dimensions through a common pool of processing-storage principles.

In TSOMs, co-activation presupposes information sharing: two IAPs are significantly co-activated if their levels of node activation largely correlate, i.e. if the same nodes respond alike to similar input words. For example, German *gemacht* tends to co-activate some of the BMUs responding to *macht*, even though *macht* appears at different points in time in the two forms (see distance matrix in Figure 6). This happens because the two IAPs for *macht* and *gemacht* share a blended pattern, which responds to two conflicting input requirements.

![Figure 6. Co-activation distances between the BMUs prompted by German *macht* and *gemacht*. Zero values mean that exactly the same nodes are responding to the same letters in the two forms. Low values indicate strong competition and high confusability between the corresponding nodes. High values indicate that levels of activation correlate poorly.](image-url)
This is symbolically represented by the word graph in Figure 7, where nodes responding to -mach- appear to be shared by several paradigmatically related verb forms. Blended patterns thus correlate with recurrent sublexical chunks. In morphologically coherent word families (e.g. inflectional or derivational paradigms), blended patterns represent perceived morphological structure.

![Figure 7](image)

**Figure 7.** A graph-based representation of a sub-paradigm of German *machen* (‘make’). Dashed lines indicate alternative connections to co-activated nodes.

In TSOMs, the syntagmatic and paradigmatic dimensions of lexical organisation interact through memory entrenchment. Large neighbour families tend to develop blended activation patterns, which are part of more IAPs and respond to the conflicting requirements of more neighbouring words. The effect is twofold. On the one hand, blended patterns provide an entrenched long-term response of the map to recurrent sub-lexical chunks (facilitation). On the other hand, blended patterns make more IAPs be simultaneously active in working memory, thereby making conflicting expectations for possible input continuations (inhibition). In the end, input frequency effects bias the map’s behaviour in either direction. Highly entropic neighbour families, where members are distributed evenly, tend to be facilitatory. However, when one family member is much more frequent than the others, the neat effect will be inhibitory, particularly on low-frequency members.

The box plot of Figure 8 (left) shows that words belonging to smaller neighbour families recruit IAPs with weaker connections. The larger the family, the stronger the connections. However, there is a significant interaction between a neighbour family size and the frequency distribution
of its members. Figure 8 (right) shows how much activation filtering it takes for a map to accurately recall a word from its IAP. Levels of filtering reflect the amount of effort needed to single out and recall the correct BMU from an IAP that contains other co-activated nodes. The more node competitors are present, the higher the level of filtering. We plotted levels of filtering for four cases: (i) low-frequency words in low-entropy families ($F_{\text{low}} \cap \text{NBH}_{\text{low}}$), (ii) low-frequency words in high-entropy families ($F_{\text{low}} \cap \text{NBH}_{\text{high}}$), (iii) high-frequency words in low-entropy families ($F_{\text{high}} \cap \text{NBH}_{\text{low}}$), and (iv) high-frequency words in high-entropy families ($F_{\text{high}} \cap \text{NBH}_{\text{high}}$). As expected, high-frequency words do not suffer from the competition of their neighbours, no matter whether they belong to a low-entropy or a high-entropy family. Low-frequency words, on the other hand, are harder to access and recall if they are surrounded by high-frequency neighbours. The problem disappears if neighbours are uniformly distributed.

Figure 8. (Left) Values of connection strength for words with different neighbour family size, ranging from 0 to more than 7 members. Cardinality of each family is given in brackets. (Right) Levels of activation filtering for four frequency classes of words: p-values show statistically significant differences between the two leftmost classes and between the two rightmost classes (see text for more detail). (Adapted from Pirrelli et al., 2014)

This dynamic reversal from facilitation to inhibition is in keeping with well-known effects of family size and frequency of neighbouring words on a variety of word processing tasks: non-word repetition (Vitevitch et
al., 1997; Vitevitch & Luce, 1998), naming (Levelt & Wheeldon, 1994), recall from verbal short term memory (Gathercole et al., 1997), phoneme identification (Pitt & McQueen, 1998). More recently, scholars have found a consistent pattern of facilitatory/inhibitory processing effects by size and frequency distribution of inflectional paradigms and inflectional classes (Baayen et al., 2006; Moscoso del Prado et al., 2004, Milin et al. 2009). Based on this evidence, we tried to assess the connection between family effects, frequency and regularity in the acquisition of verb paradigms.

Marzi and colleagues (2014) trained two different groups of TSOMs on the 50 most frequent German verb sub-paradigms (consisting of 15 forms each) for 100 learning epochs. In the first group input verb forms were administered to a map with their frequency distribution in Celex (skewed distribution). In the second group, verb forms were input five times each (uniform distribution). Figure 9 shows the time course of acquisition for the 50 verb sub-paradigms, broken down into regulars (left) and irregulars (right)\(^5\). A paradigm acquisition epoch was defined as the average time it takes for all input forms of a paradigm to be recalled correctly from their IAPs. Note that, in most cases, uniformly distributed paradigms are acquired earlier than paradigms presented with skewed distributions. A map trained on skewed distributions will tend to acquire the most frequent forms first (see Figure 3 above), thereby neglecting less frequent ones; since top-frequency words are not necessarily the most typical forms in the training set (i.e. the most similar to the majority of words), low-frequency items can not take advantage of earlier stored items, and this slows down lexical acquisition overall. However, as shown in Figure 9, the advantage between skewed and uniformly distributed paradigms has not the same impact on regulars and irregulars: a strong inverse correlation between paradigm frequency and time of acquisition holds for irregulars ($r=-.60$, $p<.0005$), but it is weaker and statistically not significant with regular paradigms. The correlation points to an important interaction between frequency and

\(^5\) We follow here the traditional descriptive assumption that regular paradigms exhibit no stem alternation. In fact, so-called irregular paradigms may differ substantially in the amount of redundancy they present. Our evidence shows that TSOMs are sensitive to levels of structural redundancy in irregular paradigms.
regularity in paradigm acquisition, based on inhibition/facilitation effects in neighbour families.

Figure 9. Time course of acquisition of regular (left) and irregular (right) German paradigms ranked by increasing learning epoch under skewed (grey circles) and uniform (white circles) learning regimes. For each paradigm, its cumulative frequency is reported in brackets. (Adapted from Marzi et al., 2014)

4.4. Regularity-by-frequency interaction

Considerable evidence has accrued on the fundamental role of paradigmatic relations as principles of non-linear organisation of word forms memorised in the speaker’s mental lexicon, facilitating their retention, accessibility and use (Carstairs & Stemberger, 1988; Baayen et al., 1997; Orsolini & Marslen-Wilson, 1997; Bybee & Slobin, 1982; Bybee & Moder; 1983; Orsolini et al., 1998; Milin et al. 2009). In particular, children have been shown to be sensitive to sub-regularities holding among paradigm cells (see, among others, Orsolini et al., 1998; Colombo et al., 2004 on Italian; Dabrowska, 2004, 2005 on Polish; Labelle & Morris, 2011 on French), in line with
“Words and Paradigms” approaches to morphological competence (see Blevins, 2006). According to these approaches, mastering the inflectional system of a language amounts to acquiring an increasing number of paradigmatic constraints on how paradigm cells should be filled in (or Cell Filling Problem: see Ackerman et al., 2009; Finkel & Stump, 2007; Pirrelli & Battista, 2000; Matthews, 1991; among others). In child acquisition of inflectional morphology, paradigm regularity is known to modulate distributional effects based on frequency (Bittner et al. 2003).

The experimental evidence of Figure 9 above shows that these effects can be replicated by TSOMs (Marzi et al., 2014). If the lexicon were a closed repository of unrelated exceptions, selective specialisation, whereby entrenched patterns are apportioned to more frequent words, would be the optimal strategy for lexical acquisition. However, we know that, in all world languages, the lexicon is a mixed bag of both morphologically irregular and regular verb paradigms. Unlike irregular paradigms, where different stems are used, more or less predictably, in different inflected forms, in regular verb paradigms, the vast majority of inflected forms of the same verb share the same stem. In this case, spreading activation and pattern blending can take advantage of redundancies in the data and use them to acquire novel combinations of existing data, irrespective of their distributions.

This two-fold, opportunistic strategy is illustrated in Figure 9, where a few highly irregular top-frequency paradigms (e.g. sein, werden, haben, können, sehen, wollen) consistently benefit from high frequencies and are learned more quickly when they are input with Celex-like distributions. In fact, high-frequency words develop deeply entrenched, tightly connected activation patterns, which are individually acquired at very early stages. However, they share very little of their information with other patterns, preventing other verb forms from taking advantage of this information through shared connections (see Figure 5.b). This strategy is advantageous in the acquisition of highly irregular paradigms, where most members are isolated and there is little if any co-activation with other forms.

On the other hand, more regular paradigms are less sensitive to distributional differences, since the map can acquire low-frequency forms by inferring them from more frequently attested forms of the same paradigm (intra-paradigmatic generalisation). This is a neighbour family
size effect (see section 4.3). Since an input word causes a distributed activation pattern over map nodes that keep memory of other previous activations, the inflected form of a regular paradigm prompts an integrated pattern where nodes responding to other inflected forms are co-activated. Co-activated nodes form a blended pattern, which responds to other non-target words, both within the same paradigm and across paradigms, and can benefit from their cumulative frequency in training. This causes a possibly noisy representation of the word, and competition with concurrent activation patterns associatively related to the current input. However, facilitatory competition turns into inhibitory competition if some members of a regular paradigm are considerably more frequent than others. The role of frequency in processing is in fact mediated by memorisation of activation patterns. Through repeated exposure to the same input word, the map gradually increases levels of activation of the target representation, by strengthening the connections between successfully responding nodes. The more frequent an input word is, the stronger its resulting activation pattern will be. If levels of activation of competing patterns are relatively balanced, adjustment of connections is quick and effective. Connection adjustment takes longer if it has to overcome the competition of one or a few stronger neighbours.

Figure 10 shows effects of the frequency-by-regularity interaction in some detail by comparing the acquisition pace of three German (sub-)paradigms out of the full set of fifty paradigms (see Figure 9), machen, finden and sein (respectively ‘make’, ‘find’ and ‘be’), which exhibit different degrees of stem regularity. machen is the prototypical example of a regular paradigm and presents no stem alternation; sein is highly irregular, and exhibits extensive stem suppletion; finden presents three stem alternants, that are shown in training with comparable frequencies. In Figure 10, for each form, a grey circle marks the time of acquisition in the Celex-like distribution, and a white circle the time of acquisition in a uniform distribution.

We observe that forms of irregular paradigms (sein and finden) take considerable advantage from skewed distributions, whereas for forms of a regular paradigm (e.g. machen) the advantage is reduced (and in many cases reversed), and time of acquisition shows no correlation with frequency effects. In particular, the time span between the first and the last word
forms to be acquired in the same paradigm is substantially short for regular paradigms. The evidence of a nearly instantaneous acquisition of all words

![Diagram](image)

**Figure 10.** The time course of word acquisition in three German verb sub-paradigms (*machen*, *finden*, *sein*) under two training regimes: realistic distributions (grey circles) and uniform distributions (white circles). Forms are ordered by increasing token frequency in the Celex-like distribution (reported in brackets).

In a regular paradigm bears witness to the ability of a TSOM to generalise already stored activation patterns to new input forms.

In the end, our evidence is in keeping with a model of dynamic acquisition where storage and processing are mutually implicated. Adaptive memories, such as TSOMs, are good at processing regular paradigms just because regular forms easily meet global expectations developed on the basis of already stored forms. At the same time, item-based memorisation is resorted to when local expectations are not supported by other members of the same family, as is the case with irregular paradigms. No computational architecture in which storage and processing are modelled as distinct components could replicate this dynamic effect in acquisition.

From this perspective, different processing strategies should reflect different storage dynamics, since systematic, structural relations between input word forms are cached as overlapping, largely automatized activation patterns, whose concurrent memorisation in the mental lexicon accounts for frequency-based effects in lexical competition, entrenchment and
5. General Discussion

Over the last decades, considerable progress in understanding the mechanisms governing lexical storage, access, acquisition and processing has questioned traditional models of language architecture and word usage based on the hypothesis of a direct correspondence (Ullman, 2004; Clahsen, 2006) between modular components of grammar competence (lexicon vs. rules), processing correlates (memory vs. computation) and their neuro-anatomical localisations in the brain left hemisphere (prefrontal vs. temporo-parietal perisylvian areas). A new type of indirect correspondence appears to suggest itself, based on a considerably more distributed and integrative view, whereby words in the mental lexicon are emergent properties of the functional interaction between different brain regions. Some of these regions are known to be associated with specific processes and structures (e.g. sensory and motor areas), while some others provide the neuro-anatomical substrate for more domain-general processing micro-functions (such as syntagmatic activation, paradigmatic co-activation, competition, storage and selection), which participate in multiple high-level linguistic functions such as phonological and orthographic processing.

A key insight of this integrative view is that different high-level effects of the operation of a common pool of processes are a by-product of the specific nature of input representations (be they acoustic or visual input signals, or sensory-motor patterns), and the inherent requirements of the processing task (e.g. visual vs. acoustic word recognition, word production, word recall etc.).

We argue here that neuro-computational models can support such a novel, bottom-up view of the language architecture. According to this view, higher-level modules, such as the mental lexicon, result from the integration of Hebbian principles of correlative learning, a pool of low-level functional processes governing storage and processing of symbolic time-series, and modality-specific levels of lexical representation. We showed that TSOMs have the potential to provide qualitative and quantitative correlates of human language behaviour, and can offer detailed explanatory insights into
the functional neurophysiology of the mental lexicon.

The TSOM-based lexical architecture described in the previous pages is intended to define a general-purpose access level of lexical information, whose primary input is the most peripheral encoding of lexical forms. No information about the lexico-semantic content of word forms or about their associated morpho-syntactic features is provided. Emerging lexical structures are set on by purely formal redundancy in surface input data, and lexical organisation is grounded on memory-based processing strategies, where input factors (e.g. word frequency, paradigm (ir)regularity, length, perceptual salience, wordlikeness) dynamically affect memorisation/acquisition and emergence of morphological knowledge. By choosing such a relatively abstract model, we wanted to show that a wide range of nontrivial observations follow from the interactive dynamics of principles of Hebbian learning and serial cognition.

The next step in the same direction should consider a network of multi-level, parallel-processing, co-organising maps synchronised through Hebbian connections. More maps can be made dynamically co-organise lexical information by associating, through Hebbian connections, lexico-semantic and morpho-syntactic information with knowledge of the differently inflected forms spelling out that information on the word-form map. In this way, we can reasonably simulate effects of inhibitory competition between concurrently activated lexical entries, and selective activation of nodes on the word-form map, as a result of pre-activation of the lexical and morpho-syntactic maps for word production.

Following Westermann and Miranda’s (2004) model of babbling, in an extended TSOM architecture each map is specialised for supporting the organisation and processing of a domain-specific level of linguistic information (e.g. printed letters, sounds, and motor parameters, morpho-syntactic and lexico-semantic features), mimicking the behaviour of specialised (unimodal) cortical areas in the human brain. Westermann and Miranda illustrate the case of a self-organising auditory map (for input sounds) coupled with a self-organising articulatory-motor map (whose input consists of motor parameters representing positions of different articulators). Thanks to Hebbian connections linking the two maps, the articulatory map can adaptively modify its internalised organisation of motor parameters by
receiving sound feedback from the auditory map. In turn, the auditory map can dynamically adjust its response as a function of the interaction with two levels of input connections: one conveying the map’s domain-specific input signal (i.e. a specific sound), the other conveying the activation-based response of the motor map in producing the same sound. In other words, the auditory map can modify its own perception of a sound, on the basis of information of how the sound is articulated by the motor map. This suggests that internalised representations of unimodal data are eventually based on more levels of information feeding back to the map and modifying its perceptual response.

A more fine-grained architecture would require implementation of a more complex network (see Pulvermüller & Garagnani, 2014), in line with recent acquisitions on the neuro-functional architecture of the language processor (e.g. Catani et al., 2005; Shalom & Poeppel, 2008; Friederici, 2012) together with advances in understanding the neuro-anatomical bases of working memory (Wilson, 2001; D’Esposito, 2007; Ma et al., 2014).

As a general point, map co-organisation through Hebbian connections can enforce multi-level integration of distributed information, whereby knowledge organised and structured according to autonomous, domain-specific principles and representations on a map, is selectively modulated through adjustment (namely enhancement or suppression) of patterns of activation that are generated by other maps (Figure 11).

Such an integrative network of (Hebbian) interconnected maps can account for the emergence of many parallel levels of structure embedding in the mental lexicon. For example, simultaneous activation of nodes corresponding to a lexical exponent and morpho-syntactic representations (feature values of person, number and tense) can sample activation patterns corresponding to fully inflected forms. All these levels of (sub-)lexical structure do not have to be hierarchically arranged in a cascade. Rather they may reflect independent, orthogonal and emergent principles of lexical self-organisation, prompted by independently associated maps, as argued for by non-derivational models of parallel language processing such as constraint-based stochastic parsing (Seidenberg & MacDonald, 1999; MacWhinney, 2004) or Sadock’s *Autolexical Syntax* (1991).

In this perspective, by tracing the developmental trajectory of lexical
storage and co-organisation on various representation levels, we can arrive at a better understanding of lexical acquisition as an adaptive multi-factorial process, and provide an explanatory, integrative account of linguistic, behavioural and neuro-physiological evidence on the mental lexicon.

Figure 11. A logical schema of the functional interaction between parallel-processing TSOMs in a multi-level lexical architecture. Maps are synchronised through two-way Hebbian connections to co-organise patterns of activation by taking input from other maps. Each TSOM develops an independent level of domain-specific organisation, but each such level is selectively modulated through enhancement and suppression patterns generated by other maps. By way of illustration, Hebbian connections from lexico-semantic, morpho-syntactic (MS), visual and motor maps to the activation pattern for the string \textit{pops} in the word-form map allow for levels of structure embedding to emerge. For simplicity, only some between-map Hebbian connections are drawn.

Acknowledgments

The present methodological reanalysis extensively draws on joint work with Marcello Ferro. We are grateful to Petar Milin and two anonymous reviewers for their precious comments and suggestions on an early version.
References


A Neuro-Computational Approach to Understanding the Mental Lexicon

A Neuro-Computational Approach to Understanding the Mental Lexicon


Levelt, W.J., & Wheeldon, L. 1994. Do speakers have access to a mental syllabary?. *Cognition* 50(1), 239-269.


