MORPHOLOGY, MEMORY AND THE MENTAL LEXICON

Marcello Ferro Giovanni Pezzulo Vito Pirrelli

Abstract: Recent experimental evidence on morphological learning and processing has prompted a less deterministic and modular view of the interaction between stored word knowledge and on-line processing. Storing a word in the mental lexicon does not simply entail keeping a faithful memory image of that word in the most compact way. It also requires encoding and manipulating such image through topological structures that are optimally adapted to word production and comprehension. Temporal Self-Organizing Maps (THSOMs) are a novel model of artificial neural network that keeps time serial information through predictive activation chains of receptors encoding both spatial and temporal information of input stimuli. The impact of this model on issues of lexical organization and morphological processing is investigated in detail through a series of simulations shedding light on the dynamics between short-term memory (activation), long-term memory (learning) and morphological organization of stored word forms (topology).

Keywords: Mental lexicon, morphological structure, word learning, Self-Organizing Maps, memory, language architecture.

1. INTRODUCTION

In classical "dual-route" models of word processing and learning (Prasada & Pinker, 1993; Pinker & Prince, 1988; Pinker & Ullman, 2002, among others), lexical roots and affixes are taken to be the basic building blocks of morphological competence, on the assumption that the mental lexicon is largely "redundancy-free". The speaker, having identified and stored the constituent parts of a word form, proceeds to discard the original word from the lexicon. The form is eventually produced by accessing and reassembling its parts. The hypothesis endorses a "direct functional correspondence" between principles of grammar organization (the lexicon *vs.* rules dichotomy), processing correlates (storage *vs.* computation) and localization of the cortical areas functionally involved in word processing (temporo-parietal *vs.* frontal areas in the human cortex, see Ullman, 2004).

It has been observed (Baayen, 2007) that a direct correspondence hypothesis, arguably the most straightforward model of the grammar-processing relation (Miller & Chomsky, 1963; Clahsen, 2006), reflects an outdated view of lexical storage as more ‘costly’ than computational operations. Alternative theoretical models put forward a more nuanced “indirect correspondence hypothesis”, based on the emergence of morphological regularities from independent principles of lexical organization (Corbett & Fraser, 1993; Wunderlich, 1996; Dressler *et al.*, 2006), whereby fully inflected forms are redundantly stored and mutually related through entailment lexical relations (Matthews, 1991; Pirrelli, 2000; Burzio, 2004; Blevins, 2006). This view prompts a different computational metaphor than traditional dual-route models: a speaker’s lexical knowledge corresponds more to one large relational database than to a general-purpose automaton augmented with lexical storage (Blevins, 2006), thus supporting a “one-route model” of word competence.

Over the past three decades, the psycholinguistic literature has provided a large body of empirical evidence intended to test the implications of dual-route and one-route models of the mental lexicon. Data, however, have so far failed to provide conclusive support to either account. Sub-word constituents are shown to play a crucial role in the processing and representation of morphologically complex words (see McQueen & Cutler, 1998 and Clahsen, 1999 for overviews). In lexical decision tasks (Taft, 1979; Whaley, 1978; Balota, 1994, for a review), target lexical bases are effectively primed by earlier presentation of regularly-inflected related forms (*walked* → *walk*), but they are not primed by irregular inflections (e.g. *brought vs. bring*). The effect is interpreted as showing that *walked* activates two distinct lexical representations, one for the stem *walk* and the other for the affix *-ed*. One-route models of morphological processing, on the other hand, account for dissociation effects of this kind in terms of type/token frequency factors, phonological and semantic similarity, or both (e.g. Eddington, 2002; Ellis & Schmid, 1998; Joanisse & Seidenberg, 1999; Rueckl & Raveh, 1999).

Recent empirical findings suggest that surface word relations constitute a fundamental domain of morphological competence, with particular emphasis on the interplay between “form frequency”, “family frequency” and “family size” effects within morphologically-based word families such as inflectional paradigms (Baayen, Dijkstra & Schreuder, 1997; Taft, 1979; Hay, 2001; Ford, Marslen-Wilson & Davis, 2003; Lüdeling & De Jong, 2002; Moscoso del Prado Martín *et al.*, 2004). Research on speech errors (Stemberger & Middleton, 2003) suggests that English present and past tense forms are in competition, and that this competition is modulated by the a-priori probabilities of the vowels in these forms, even if they are regular (Tabak *et al.*, 2005). Maratsos (2000) reports that many individual verbs are used by children in both correct and over-generalized forms (e.g.

brought and **bringed*) for a long period, thus supporting a more dynamic, frequency-based competition between regular and irregular forms than dual-route accounts are prepared to acknowledge. The assumption that both regular and irregular forms are stored in the lexicon seems to go a longer way towards a competition-based account.

That more than just storage is involved, however, is suggested by interference effects between false morphological friends (or opaque pseudo-derivations) such as *broth* and *brother*, which share a conspicuous word onset but are not related morphologically (Longtin *et al*., 2003; Rastle *et al*., 2004). These and other similar results, observed particularly but not exclusively for Semitic languages (see Frost *et al*., 1997 and more recently Post *et al*., 2008), show that as soon as a given letter sequence is fully decomposable into morphological formatives, word parsing takes place automatically, prior to (or concurrently with) lexical look-up.

To sum up, both associative and dual-mechanism models find it hard to account for the entire body of presently available psycholinguistic evidence on word learning and processing. All in all, the evidence appears to point to a less deterministic and modular view of the interaction between stored word knowledge and on-line processing than dual-route approaches are ready to acknowledge. For example, if lexical look-up is the first step in word processing, then pseudo-affixed monomorphemic words such as *brother* should not undergo decompositional processing because they are readily found in the lexicon. As we saw, however, this is contrary to evidence on automatic processing. On the other hand, there is no way to account for such effects in terms of either variegated analogy (of the sort used by example-based approaches) or phonological complexity and perceptual subtlety of the input word (as suggested by McClelland & Patterson, 2002). Analogies and inflectional rhyming patterns have to exhibit a clear morphological status. In current connectionist thinking, however, such status is taken to be merely epiphenomenal.

The currently emerging view sees word processing as the outcome of simultaneously activating patterns of cortical connectivity reflecting (possibly redundant) distributional regularities in input data at the phonological, morpho-syntactic and morpho-semantic levels. At the same time, there is evidence to argue for a more complex and differentiated neuro-biological substrate for human language than connectionist one-route models can posit (Post *et al*., 2008), suggesting that brain areas devoted to language processing maximize the opportunity of using both general and specific information simultaneously (Libben, 2006), rather than maximize processing efficiency and economy of storage.

To our knowledge, no current computational model of word learning can account for such a complex interaction. Both symbolic and connectionist approaches have so far laid exclusive emphasis on processing aspects of word competence only, whereby morphological productivity is modelled as the task of outputting a (possibly) unknown word form (say a novel inflected form) by taking as input its corresponding lexical base. Such a "derivational" approach to word competence (Baayen, 2007) ends up obscuring the deep interplay between storage and computation, laying emphasis on a merely procedural view of morphological competence as the "ability to play a word game". In classical connectionist architectures (Rumelhart & McClelland, 1986), the internal organization of inflected forms in the mental lexicon is modelled by the pattern of connections between the hidden layer and the output layer in a multilayered perceptron trained on mapping lexical bases onto inflected forms (e.g. *go* → *went*). The resulting lexical organization appears to be contingent upon the requirements of the task of generating novel forms. In principle, different tasks may impose different structures on the mental lexicon.

In this paper, we shall take a somewhat different approach to the problem. In line with the psycholinguistic evidence reviewed above, we assume that word storage plays a fundamental role in both word learning and processing. The way words are structured in our long-term memory (the mental lexicon) is key to understanding the mechanisms governing word processing and productivity. This perspective offers a few advantages. First, it allows scholars to properly focus on word productivity (the *explanandum*) as the by-product of more basic memory strategies (our *explanans*) that must independently be assumed anyway to account for fundamental aspects of word learning (including but not limited to memorization of word forms). Secondly, it opens up new promising avenues of scientific inquiry by tapping the large body of empirical evidence on short-term and long-term memorization strategies for serial order (see Baddley, 2007 for a comprehensive overview). Furthermore, it gives the opportunity of using sophisticated computational models of language-independent memory processes (see Botvinick & Plaut, 2006; Brown Preece & Hulme, 2000; Henson, 1998; Burgess & Hitch, 1996, among others) to shed light on language-specific aspects of word encoding. Finally, it promises to provide a comprehensive picture of the complex dynamics between computation and memory underlying morphological processing as portrayed by the psycho-linguistic and neuro-linguistic literature on the topic. Put in a nutshell, the processing of unknown words requires mastering rule-governed combinatorial processes. In turn, these processes presuppose knowledge of the sub-word units to be combined. We argue that preliminary identification of the basic inventory of such units depends on memorization of their complex combinations. As we shall see in more detail later in the paper, current models of memory for serial order assume a much deeper interaction between stored knowledge and on-line processing. The way information is stored reflects the way such information is dynamically represented, and eventually accessed and retrieved as patterns of concurrent activation of memory areas. According to the view endorsed here, memory can "both hold information and manipulate it". This is a far cry from traditional computational devices for storage allocation and serial order representation (such as ordered sets, strings and the like) which provide built-in means of serializing order information through chains of pointers which are accessed and manipulated by independently required recursive algorithms.

The paper is structured as follows. We first offer an overview of issues of memory for serial order (sections 2 and 3) to then move on to their possible connections with aspects of parallel lexical processing (section 4). A few computer simulations with a novel family of Kohonen's Self-Organizing Maps (SOMs) are then shown in section 5, followed by a general discussion and some concluding remarks (sections 6 and 7).

2. MEMORY FOR SERIAL ORDER

A fundamental characteristic of the human language faculty is the ability to retain sequences of items (*e.g.* letters, syllables, morphemes or words) in the so-called "Short Term Memory" (STM). Since Baddeley & Hitch's pioneering work, the processing mechanisms underlying human STM have been understood within an influential model of working memory based on the so-called phonological loop (Baddeley & Hitch, 1974; Baddeley, 1986, 2006). The model is assumed to contain three separate but interacting components of limited capacity: a "central executive"*,* responsible for attention and control processes, and two slave buffer stores, one specialized for containing visuo-spatial information, the other for containing verbal information. In particular, the verbal buffer comprises a temporary phonological store in which auditory memory traces decay over a period of few seconds, unless they are refreshed through repeated (either overt or covert) vocalization. The mechanism, known as the "phonological loop", reflects the rather common experience of repeated articulation of a telephone number for long enough to be able to dial it, and explains a variety of behavioural facts about repetition of word sequences (Baddeley, 1966, 1974), including defectiveness of the short-term phonological store in patients with specific central articulatory deficits (aphasic patients with dyspraxia, Waters, 1992). Over the last few years, there has been growing recognition of the deep connection between STM and language processing (Baddeley, 2003). Lexicon and grammar appear to be strongly interconnected in this perspective, as they are both "Long-Term Memory" (LTM) containers.

2.1 Factors affecting STM

2.1.1 Length effects

Human subjects are reported to accurately recall up to “five/six arbitrary items” in a one-off sequence. However, they find it increasingly hard to accomplish the same task as soon as the length of a sequence exceeds the five units. Since Miller’s (1956) pioneering work, scholars have investigated the growing difficulty of immediately recalling longer arbitrary sequences in terms of limits in the STM “storage capacity”. Storage capacity defines the maximum number of memory traces whose activation can concurrently be sustained through a short time interval with no rehearsal and no support of long term knowledge (see Cowan, 2001, for an overview). When performance is suboptimal, accuracy in item recall is distributed across the sequence unevenly, with a characteristic U-shaped pattern: early items happen to be reported more accurately than later items (“primacy gradient”), with final items being less prone to errors than items in the middle (“recency gradient”).

2.1.2 Transpositions and inherent similarity effects

The most common errors in serial recall are order errors or “transpositions”. An aspect of these errors is their distribution in the sequence: erroneous items are mostly recalled in a position that clusters around their correct position, rather than being randomly distributed. For example the string *trap* is often mistyped as *tarp*, where at the point where an *r* should be produced an upcoming *a* is produced instead, to be eventually suppressed at the next stage, replaced by the missing *r*. This provides evidence that upcoming responses are already active before the point in time at which they are produced, thereby accounting for co-articulation effects in serial recall, in which production of the current item in the sequence is affected by anticipation of the upcoming item. Consequences of parallel competing activations of consecutive items are also witnessed in so-called “similarity” effects upon recall: serial recall of lists of similar items is considerably worse than recall of lists of dissimilar items. Accordingly, recall of the sequence <*trap*, *crap*, *dark*> should be more difficult than recall of the list <*trap*, *dog*, *smell*>, due to the greater confusability of items in the first list.

2.1.3 Grouping

Another important factor that appears to influence human performance in immediate serial recall is provided by the temporal rhythm with which verbal items are presented. Performance is enhanced by presenting items with a specific temporal grouping, as opposed to an evenly-spaced presentation (Hebb, 1961). The incidence of order errors in recall decreases between items that occupy different within-group positions (Frankish, 1985), thereby pointing to a notion of inter-item confusability that rests on distributional (as opposed to phonological or merely formal) similarity. For example, phonological segments that form the onset of adjacent syllables tend to be recalled in reversed order, as in classical slips of the tongue like *car park* → *par cark*. A second type of positional errors is found between trials. An erroneous item in one recalled sequence is more likely than chance to have occurred at the same position in the previous sequence (Conrad, 1960; Estes, 1991). In the STM literature, this evidence is typically interpreted as supporting the idea of "disjunctive memory representations" for list items (their content) and for their position in the list (their context). In fact, confusion between contextually similar items is a hallmark of linguistic structure. Units that occupy the same position and are mutually substitutable in context tend to be classified as instances of the same type and are hence more likely confused. Grouping effects on serial recall thus shed light on the relationship between serial recall and structure.

2.1.4 Chunking

Serial sequences are known to be recalled more easily if they are repeatedly encountered in the subject's input. If asked to recall a sequence of random words, subjects begin to make errors once the sequence is longer than six. If the words are concatenated in a meaningful sentence, however, then a span of 16 or more words is recalled correctly (Baddeley, 2000). This shows that subjects use long-term information to integrate the constituent words into a smaller number of "chunks" (Miller, 1956). Chunk integration causes the STM capacity (or span) to be set by the number of chunks, rather than the number of items.

2.2 STM, LTM and the mental lexicon

Aspects of the phonological loop underlying verbal short-term memory also underlie "vocabulary acquisition". Gathercole & Baddeley (1989) describe a group of children with Specific Language Impairment (SLI), whose limited capacity of repeating a sequence of words was found to be highly correlated with a strongly impaired ability to repeat "nonword items" and a comparatively impoverished vocabulary. Such a deep interconnection between immediate serial recall, nonword repetition and vocabulary acquisition has been confirmed by a large array of empirical and experimental evidence on both adults and infants (Papagno *et al*., 1991; Service, 1992; Shallice & Vallar, 1990, to mention a few). Lexical acquisition hence requires the full capacity of retaining temporal sequences of items.

Various aspects of lexical storage are accountable in terms of basic mechanisms of memory for serial order. Lexical items are known to be stored "in waves", with confusable items usually having similar beginnings and similar endings (e.g. *anecdote vs. antidote*, *musician vs. magician*), with initial sounds being more confusable in short words and final sounds being more confusable in longer words. The phenomenon, known after Aitchison (1994) as the "bathtub effect", is not just due to selective attention, but appears to reflect primacy and recency memory gradients familiar from the literature on STM. Despite extensive patterns of redundant morphological structure, chunking is ubiquitous in the mental lexicon. Contrary to traditional wisdom in dual-route approaches, storage of morphologically complex full forms is not restricted to irregular words (but see Pinker & Ulman, 2002 for qualifications of this point). Regular word forms may also leave whole memory traces in the mental lexicon if their frequency falls above a certain threshold (6 per million, according to Alegre & Gordon, 1999, but see De Vaan *et al.,* 2007 for different estimates). Typically, a word frequency effect goes hand in hand with the absence of both stem and affix frequency effects, supporting the idea that whole word entrenchment blurs morphological structure, with the whole taking precedence over its parts. This is not to deny the existence and functional role of morphological structure in the mental lexicon. As observed by Hay & Baayen (2005), stems and affixes may well develop their own lexical representations. Nonetheless, such representations crucially depend, for their existence, on the continuing degree of probabilistic support received from the network of long-term associative paradigmatic relations.

We may wonder about the possible advantages that chunking offers for lexical processing. One of them is "predictive selection" of the most likely input sequences, with consequent probability-driven elimination of unlikely (but possible) segmentations. Experimental studies based on event-related potentials and eye-movement evidence, for example, show that people use prior (lexical and semantic) contextual knowledge to anticipate upcoming words (Altmann & Kamide, 1999; Federmeier, 2007). DeLong *et al*. (2005) demonstrate that expected words are pre-activated in the brain in a graded fashion, reflecting their expected probability. This provides a solid empirical ground to probabilistic approaches to lexical prediction and gaze planning. Ferro *et al*. (2010) offer a computational model of the interlocked relationship between processes of lexical self-organization and active sensing strategies for reading that exploit expectations on stored lexical representations to drive gaze planning. This can explain why the capacity to repeat non words is a good predictor of whether or not the child is likely to encounter reading problems (Baddeley & Gathercole, 1992; Gathercole & Pickering, 2001).

Predictive selection is not the only advantage offered by chunking. The STM literature shows that storage may play a fundamental role in "language processing". Since a chunk takes one store unit of the short-term span irrespectively of length, chunking augments the capacity of the STM system to maintain and manipulate longer and more complex input sequences. In principle, the process is unbounded. 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 organization of the input stream.

Given the combined evidence reviewed here, two general points can be made. First, the strong interaction between long-term and short-term storage appears to put a premium on "context-sensitive redundancy", "fluency" and "chunking", while penalizing autonomy of chunk-internal units, as they crucially undermine the beneficial effect of chunking on the STM span. Secondly, given a level of chunking, not all chunks are equally frequent. Some chunk-internal units are more frequent than others and some within-chunk transitions from one unit to its successors are more predictable than others. This sheds light on another important fact about chunking in lexical storage: the inherent probabilistic gradedness of lexical structure (Hay & Baayen, 2005).

3. COMPUTATIONAL MODELLING OF SERIAL MEMORIES

How do we manage to recall and repeat sequences of serially-ordered items? What kind of available brain memory structures and internalized representations provide support to the ordered activation of items in time? Obtaining serial recall from a biologically inspired, parallel system is a far from trivial task. Three basic mechanisms have been suggested in the computational literature to deal with the storage and retrieval of serial order: i) item chaining, ii) time-bound parallel activation of competing items, and iii) association of items with positional slots. The models making use of each such mechanism are known in the literature respectively as "chaining models", "ordinal models" and "positional models".

3.1 Chaining models

Some of the earliest psychological accounts of serial order postulate that action sequences are represented as chains made up of unidirectional stimulus‒response links. The simplest chaining models assume only pairwise associations between adjacent elements of a sequence (e.g. Wickelgren, 1965) and cues that consist of the preceding response only. This is equivalent to Markov first-order models, where the probability of having 'C' in the sequence 'ABC' is entirely determined by the conditional probability of finding 'C' given 'B'. Higher-order Markov models can be construed. For example, the probability of finding 'C' in the sequence 'ABC' can be conditional on the probability of finding both 'A' and 'B' preceding it. This can be scaled up to higher order contexts, assuming remote associations as well as adjacent ones (e.g., Ebbinghaus, 1964; Slamecka, 1985; Elman, 1990; Jordan, 1986).

Criticism of chaining models goes back to pioneering work by Lashley in the 50's (see Lashley, 1951; Houghton & Hartley, 1995, for an extensive review). It is commonly pointed out (most recently by Henson, 1998) that chaining models are token-based and that they typically face the problem of distinguishing repeated elements in a sequence. For example, in order to represent a word like '#EVERY#' as a sequence of associative links between characters, 'E' must be linked to both 'V' and 'R'. Hence, in recalling the word '#EVERY#' by going through a chain of links, it is not clear which item should follow the first instance of 'E'. Higher order models overcome this problem, but only at the expense of using distinct token representations (e.g. the bigrams '#E' and 'VE') as instances of the same type 'E'. We shall return later to this style of symbol representation, known in the literature as "conjunctive coding", in Section 4 below.

3.2 Ordinal models

One of the basic insights of Lashley's seminal work was that immediate recall of the order of appearance in a sequence of input stimuli requires their parallel activation and a conflict resolution mechanism governing their mutual competition. All input stimuli call for selection through their background activation, like customers at a crowded bar trying to attract the attention of a single bartender. Customers are still served one at a time but no ordered structure exists.

Ordinal models reflect this basic insight. Elements are encoded along a single dimension representing their strength of activation. Order is then defined by the relative values on that dimension. Grossberg (1978) assumes that order is stored in a primacy gradient of strengths, such that each element is stronger than its successor. In so-called "competitive queuing" models (Grossberg, 1978; Houghton, 1990) the order of elements can be retrieved by the iterative process of selecting the strongest element and then suppressing it (Page & Norris, 1998). Unlike chaining, ordinal models are effective in explaining transposition errors. Random oscillations in levels of item-wise activation make adjacent items more likely to be transposed, as their activation levels are closer and thus more likely to be confused in retrieval. Yet, ordinal models are subject to problems of item repetition in the same sequence, as in the example of '#EVERY#' mentioned above.

3.3 Positional models

In positional models, items are associated with memory slots and retrieved by accessing their separate addresses in memory. Many positional models exist that try to specify the exact nature of the positional codes and their mathematical and psycho-physiological correlates (see Brown Preece & Hulme, 2000; Henson, 1998; Burgess & Hitch, 1996, among others). In all of them, each occurrence of an item is assumed to create a new token in short-term memory (as in multiple-trace theories, e.g., Hintzman, 1986). These tokens are episodic records that a particular item occurred in a particular spatiotemporal context. Hence, representation of an item at the start of a sequence is quite different from the representation of the same item at the end of a sequence. STM is not viewed as a subset of active long-term memory representations (Cowan, 1993), but as a set of new, episodic tokens, thus allowing for representation of sequences with repeated items (Henson, 1996).

There is, however, at least one area that has proven challenging for positional models. This involves cases where serial recall is influenced by long-term, or background knowledge about sequential structure (see Baddeley, 1964 among others). Strings of letters, for example, are found to be better recalled if adjacent items are also likely to be sequenced together in existing words of a language. In these examples, short-term memory for serial order is seen to depend on background knowledge concerning domain-specific regularities in sequential structure. Note that in this and similar cases, recall for highly probable sequences is better than for less probable ones. Hence, the relevant background knowledge involves transition probabilities among specific items. It is this that makes the observed effects difficult for context-based models to address.

4. ISSUES IN PARALLEL LEXICAL PROCESSING

Long-term storage of words depends critically on phonological short-term memory processes. A number of further constraints on modelling lexical storage follow from this premise. Arguably, the most fundamental such constraint is that lexical representation, organization and access must be based on “parallel processing systems” that realistically mirror brain functional processes. This opens up new, challenging perspectives on basic issues underlying lexical architectures and their role and position in the language edifice.

The vast literature on STM and LTM processes has had the unquestionable merit of throwing in sharp relief some fundamental issues concerning the representation and manipulation of time-bound constraints over symbolic sequences. Word forms are primarily strings of sounds or characters and so the question of their representation and acquisition in time is logically prior to any other processing issue. Perhaps with the only notable exception of connectionist models, however, coding issues have suffered unjustified neglect by the Artificial Intelligence research community over the last 30 years. On the other hand, the mainstream connectionist answer to the problem of time series coding in parallel processing systems, namely “conjunctive coding”, seems to have eluded some core issues.

As a first approximation, conjunctive coding (e.g., Coltheart *et al*., 2001; Harm & Seidenberg, 1999; McClelland & Rumelhart, 1981; Perry, Ziegler & Zorzi, 2007; Plaut *et al*., 1996) represents the word form *CAT* by activating one representational unit that stands for *C* in conjunction with the first position, another for *A* in the second position, and another for *T* in the third position. This representation has proven useful for model building and theory testing in the domain of parallel lexical processing and learning, but it looks more like a convenient way out than a principled solution.

First, conjunctive codes are typically assumed to be available in the input (or encoding) layer in the form of a built-in repertoire of context-sensitive Wickelphones, the issue of their origin/acquisition being just swept under the carpet. Now, language learning is much more than making inferences about fully-developed sign-based lexical representations of some kind. Input representations can be noisy, crucially underspecified and may develop through time. Any model of language learning that presupposes stable full-fledged sign-based representations from the outset leaves much to be explained. A second related issue is the acquisition of phonotactic knowledge. Speakers are known to exhibit differential sensitivity to diverse sound patterns. Effects of graded specialization in the discrimination of sound clusters and lexical well-formedness judgements are the typical outcome of acquiring (the phonotactics of) a particular language. If such patterns do not develop through learning but are part and parcel of the encoding layer, the same processing system cannot be used to deal with different languages exhibiting differential sound constraints.

A third limitation of conjunctive coding is that phonemes and letters are bound with their context. This means that two elements like ‘#E’ and ‘VE’ representing two instances of the same letter ‘E’ in ‘#EVERY#’ are in fact as similar (or as different) as any two other elements. We are just left with "token" representations, the notion of "type" of unit remaining out of the representational reach of the system. This makes it difficult to generalize knowledge about phonemes or letters across positions (the so-called "dispersion problem": Plaut *et al*., 1996; Whitney, 2001). It is also difficult to align positions across word forms of differing lengths (i.e., the "alignment problem": see Davis & Bowers, 2004), thus hindering recognition of both shared and different sequences between morphologically-related forms. The failure to provide a principled solution to alignment problems (Daugherty & Seidenberg, 1992; Plaut *et al*., 1996; Seidenberg & McClelland, 1989) is particularly critical from the perspective of morphology learning. Languages wildly differ in the way morphological information is sequentially encoded, ranging from suffixation to prefixation, sinaffixation, apophony, reduplication, interdigitation and combinations thereof. For example, the alignment of lexical roots in three as diverse pairs of paradigmatically related forms such as English *WALK-WALKed*, Arabic *KaTaBa-yaKTuBu*, German *SPReCHen-geSPRoCHen* requires substantially different processing strategies. Precoding any such strategy into lexical representations (e.g. through a fixed templatic structure that separates the lexical root from other morphological markers) would have the neat effect of slipping in morphological structure directly into the input, thereby making input representations dependent on languages. A far more plausible solution would be to let the processing system home in on the right sort of alignment strategy through repeated exposure to a range of language-specific families of morphologically-related words. But this is exactly what conjunctive coding cannot do.

To our knowledge, there have been three attempts to tackle the issue within a connectionist framework: "Recursive Auto-Associative Memories" (RAAM; Pollack, 1990), "Simple Recurrent Networks" (SRN; Botvinick & Plaut, 2006) and "Sequence Encoders" (Sibley *et al*., 2008). The three models set themselves different goals: i) encoding an explicitly assigned hierarchical structure for RAAM, ii) simulation of a range of behavioural facts of human Immediate Serial Recall for Botvinick & Plaut's SRNs and iii) long-term lexical entrenchment for the Sequence Encoder of Sigley *et al*. In spite of their considerable differences in mechanisms and properties of learning, all systems share the important feature of modelling storage of symbolic sequences as the by-product of an "auto-encoding" task, whereby an input sequence of arbitrary length is eventually reproduced on the output layer after being internally encoded through recursive distributed patterns of node activation on the hidden layer(s). Serial representations and memory processes are thus modelled as being contingent on the task. In particular, Botvinick & Plaut's paper makes the somewhat paradoxical suggestion that human performance on immediate serial recall develops through direct practice on the task of word repetition. Moreover, STM effects appear to be accounted for in terms of a long-term dynamics dictated by the process of weight adjustment through learning. Although LTM effects are known to increase short-term storage capacities, the developmental evidence shows that the causal relationship is in fact reversed, with children with higher order STM being able to hold on to new words for longer, thus increasing the likelihood of long-term lexical learning (Baddeley, 2007).

In the remainder of this paper we describe a novel computational architecture for lexical processing and storage. The architecture is based on Kohonen's Self-Organizing Maps (SOMs; Kohonen, 2001) augmented with first-order associative connections that encode probabilistic expectations (so called, Temporal Hebbian SOMs, or THSOMs for short; Koutnik, 2007; Pirrelli *et al*., 2010; Ferro *et al*., 2010). We shall show that THSOMs define an interesting class of general-purpose memory models for serial order, exhibiting a non-trivial interplay between STM and LTM processes. At the same time, they simulate incremental processes of topological self-organization whereby lexical sequences are arranged in minimally entropic stochastic hierarchies. We shall also discuss properties of such hierarchies that make them interesting morphological structures.

5. TEMPORAL SOMS FOR LEXICAL PROCESSING

THSOMs are storage devices consisting of grids of memory nodes with a short-term and a long-term dynamic. The "short-term dynamic" is based on parallel activation of topologically-organized memory nodes exhibiting dedicated sensitivity to stimuli that occur in specific spatio-temporal contexts. The "long-term dynamic" unfolds through training: i) nodes are made more sensitive to particular classes of stimuli; ii) inter-node Hebbian connections are attuned to transition probabilities between temporally adjacent stimuli as they are observed in the training data.

Being generic memory models for serial order, THSOMs are not designed to perform any particular task. Nonetheless, for our present concerns, they exhibit a processing behaviour that makes them similar to stochastic Markov models whose states are topologically clustered nodes and state transitions are normalized inter-node Hebbian connections. In what follows, we first provide a sketchy account of the basic dynamics of THSOMs learning sequences of symbols (e.g. digits, letters, syllables) to then move on to a series of small-scale simulations that are intended to give the reader an intuitive flavour of their behaviour.

5.1 THSOMs in action

Figure 1 illustrates the architecture of a THSOM consisting of *N* memory nodes (or “map nodes”) arranged for simplicity along one dimension. Input to the THSOM is represented by a *D*-dimensional *“*vector” X of input codes on the map’s “input layer”.

All nodes in a THSOM are linked to the input vector through weighted connections defined over the “spatial connection layer”. When a stimulus is presented to the map, encoded on the input layer, all nodes are activated in parallel. Node activation is short-term, lasting one time step, until another input stimulus is presented. The degree of activation of a map node is a function of the “correlation” between values of the input vector and weights on the spatial connections to the node. The higher the correlation, the stronger the node activation. When correlation is high, the node is said to have a faithful “memory trace” of the input vector.

Memory traces develop through training. Training takes place iteratively, upon repeated presentation of input stimuli, and consists in adjusting weights on the spatial connection layer for them to get closer to the current values on the input layer. Weight adjustment does not apply evenly across map nodes and time steps, but depends on: a) node’s correlation to the input vector, b) map’s learning rate and c) space topology. At each input presentation, the most strongly adjusted node is the most highly activated one, called “Best Matching Unit” (BMU). All other nodes are adjusted as a (Gaussian) function of i) their distance from BMU on the map (the “neighbourhood radius”), and ii) the map’s current “learning rate”. Weights of nodes that lie close to BMU are made more similar to input values than weights of nodes lying further away from BMU. After adjustment, the time counter is increased by one tick, the map’s activation is reset and another input stimulus is encoded. Both learning rate and neighbourhood radius go down through time to simulate the behaviour of a brain map losing its plasticity (Kohonen, 2001).

Unlike classical SOMs, THSOMs can also learn synchronization of two nodes that fire at consecutive time steps. A THSOM can remember, at time *t*, its state of activation at time *t-1* and can make an association between the two states. This is done on the layer of pre- and post-synaptic connections linking each single node to all other nodes on the map, called the “temporal connection layer” (see Figure 1). Synaptic weights are adjusted by Hebbian learning (Hebb, 1949): the temporal connection between two consecutively firing BMUs is potentiated, and the temporal connections between all other nodes and the current BMU are depressed.

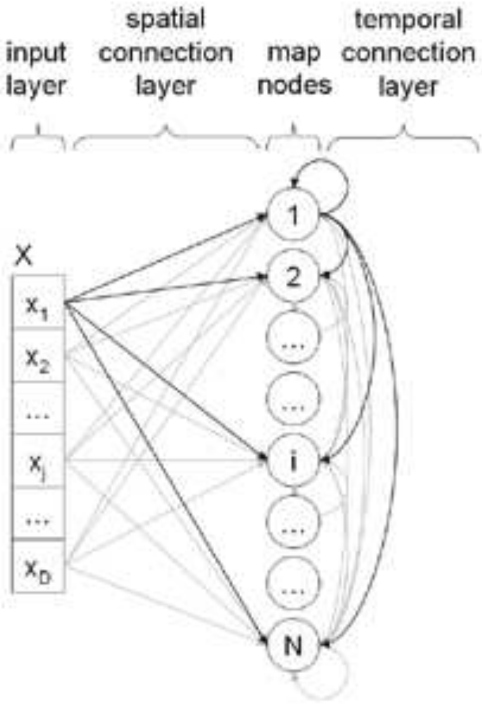


FIGURE 1. OUTLINE ARCHITECTURE OF A THSOM

In THSOMs, temporal connections affect short-term node activation. Activation of a map node at time *t* is an inverse function of the summation of two vector distances: a) the distance between the input vector and the node's spatial connection weights, and b) the distance between the node's pre-synaptic temporal weights and the whole state of activation of the map at time *t-1*. As a result, a THSOM trained on time series of input vectors develops i) a topological organization of nodes by their sensitivity to similar input vectors (or spatial similarity) and ii) a first-order time-bound correlation between BMUs activated at two consecutive time steps. Because of this second activation component, the same symbol can fire two different BMUs depending on its left context. This is due to the fact that, on the temporal layer, the two BMUs correlate with different predecessors, thereby receiving different degrees of pre-synaptic support.

Differing strategies of weight adjustment on the temporal layer can lead to considerably differing topological structures. We shall consider here two alternative regimes. Panels a) and b) of Figure 2 illustrate a "localist" temporal learning strategy originally proposed by Koutnik (2007). Potentiating (LTP) connections between consecutively activated BMUs (at times *t-1* and *t*) involve two units only (Figure 2a). Depressant (LTD) connections involve all nodes other than the BMU at time *t-1* on the one hand, and the single BMU at time *t* on the other hand (Figure 2b).

Pirrelli *et al*. (2010) propose the "distributed" activation strategy illustrated in panels c) and d) of Figure 2. Both potentiation (LTP) and depressant (LTD) connections affect a neighbouring area centred around the BMU at time *t*. By spreading the temporal support over the neighbourhood of the current BMU (as a Gaussian function of the distance from BMU), the map appears to be more prone to develop dedicated nodes for context-specific occurrences of the same letter. The ensuing simulation is intended to elaborate this point.

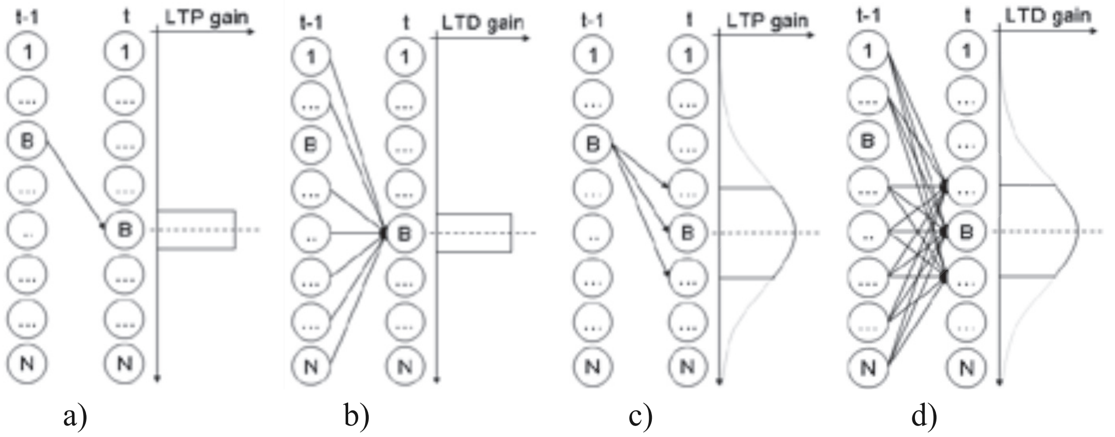


FIGURE 2. "LOCALIST" AND "DISTRIBUTED" WEIGHT ADJUSTMENT STRATEGIES

5.2 Simulation 1: hierarchical memory structures

Figure 3 shows chains of short-term activation peaks in a 7x7 bi-dimensional THSOM after presentation of the strings '#ABC', '#ACB', '#BCA', '#BAC', '#CAB', '#CBA'. Strings are shown to the map one letter at a time. The start-sequence symbol '#' is appended at the beginning of each sequence to tell the map that a new string is being shown. In the figure, each node is labelled with a letter to indicate that the node is most sensitive (above a set threshold) to that particular letter. At each input exposure, the Best Matching Unit (BMU, highlighted in bold in the figure) represents the map's highest response to the currently input letter. Solid arrows represent temporal connections linking two consecutively activated BMUs.

Note that several BMUs in the map are activated by the same letter: five BMUs for 'A', five for 'B' and five for 'C'. Each such node turns out to be specialized for a specific occurrence of the letter in a "unique left context". There is a single BMU for 'A' in initial position (in '#ABC' and '#ACB'), one for 'A' preceded by '#B' (in '#BAC'), one for 'A' preceded by '#C' (in '#CAB'), one for 'A' preceded by '#BC' (in '#BCA') and another one for 'A' preceded by '#CB' (in '#CBA').

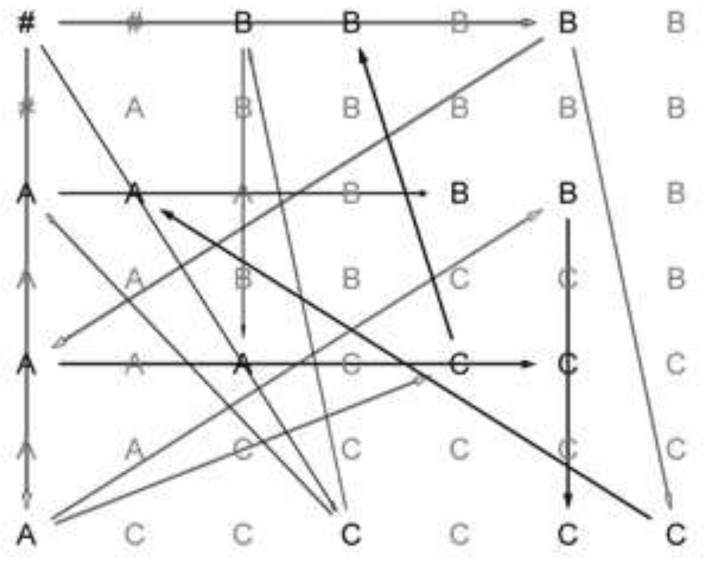


FIGURE 3. TEMPORAL ACTIVATION PATHS IN A THSOM

Context-sensitivity is something that THSOMs develop through training. The THSOM in Figure 3 was trained, over 100 sessions, on all possible permutations of 'A', 'B' and 'C'. We used the "distributed" weight adjustment strategy proposed by Pirrelli and colleagues (2010; see section 5.1 for details). After training, the THSOM has ostensibly stored all permutations as a tree-like hierarchical structure of nodes, starting with a '#' node at the root of the tree and branching out as soon as two (or more) different nodes can be alternative continuations of the same history of past activated nodes, as illustrated in Figure 4. The length of the history of past activations defines the "order of memory" of the map. It can be shown that this type of hierarchical organization maximizes the map's expectation of an upcoming symbol in the input string or, equivalently, minimizes the entropy over the set of transition probabilities from one BMU to the next. This is achieved through a profligate use of memory resources, whereby several nodes are recruited to be most sensitive to contextually specific occurrences of the "same letter".

input strings temporal paths

$$\left.\begin{array}{l} \#CBA \\ \#CAB \end{array}\right\} \Rightarrow \#\rightarrow C \begin{array}{l} \rightarrow B \rightarrow A \\ \rightarrow A \rightarrow B \end{array}$$

FIGURE 4. A TREE-LIKE LEXICAL STRUCTURE

A tree-like lexicon is not the only storage structure that a THSOM can possibly develop after training. Figure 5 shows a different organization of a THSOM trained on Koutnik's localist learning regime (section 5.1). In the figure, we note a more parsimonious use of nodes, with a maximum of 4 BMUs per letter. This means that a single node can 'recognize' several occurrences of the same letter in different contexts.

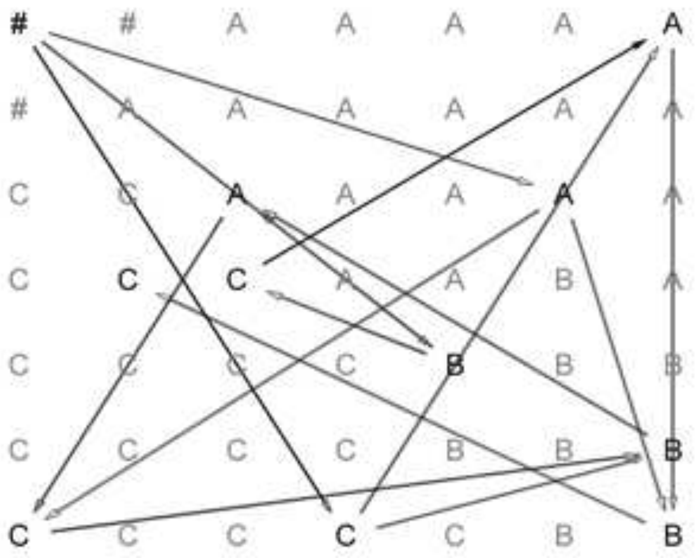


FIGURE 5. A DIFFERENT STORAGE STRUCTURE IN A THSOM

Figures 3 and 5 illustrate an exemplar case of dynamic trade-off between time and space. If the map has to devote unique BMU chains to sequences of letters, it has to recruit more nodes per symbol. Fewer specialized nodes mean shorter long-term memory spans. To be more concrete, let us take a closer look at Figure 3. Each of the two 'C' nodes at the bottom right corner of the map is preceded by a different 'B' node: a 'B' node preceded by 'A' and a 'B' node preceded by '#'. This means that the map's context-sensitivity spans over more than one letter to the left, thereby simulating a memory order greater than 1. In Koutnik's learning regime this takes place only to a limited extent. A localist temporal support creates isolated peaks of node activation (see Figure 2a). Local peaks eventually act as strong attractors for all surrounding nodes, making it very hard for higher-order memory nodes to emerge. Accordingly, a single 'C' node at the bottom left corner of Figure 5 has two pre-synaptic connections with two different 'A' nodes, one preceded by a 'B' node and the other preceded by a '#' node. Thus, the map behaves like a first order Markov chain.

5.3 Simulation 2: string prediction

In order to assess the impact of storage organization on the ability of a THSOM to predict upcoming letters, we tested the two training regimes on a "string recognition task" (Ferro *et al.,* 2010). The task consists in the map's going through familiar strings of written letters. Letters are shown one at a time from left to right. The map tries to anticipate upcoming (masked) letters on the basis of already unmasked ones. The task has several possible connections with the proactive reading strategies used by a skilful reader in scanning a written text (Ferro *et al*., 2010).

A 30x30 THSOM was trained on 66 Italian present indicative forms whose frequency distributions were sampled from the Calambrone section of the Italian CHILDES sub-corpus (MacWhinney, 2000). Two training sessions were carried out, one on a distributed learning regime, the other on its localist variant. Figure 6 reports the results of both tests in terms of overall per word accuracy of prediction and number of predicted chunks by chunk's length. The upper panel shows that the map's performance after distributed training develops considerably more and longer chunks than the same map after localist training. This impacts overall prediction accuracy on trained words, which scores 48.5% in the first test (distributed learning), and 37.9 in the second test (localist learning).

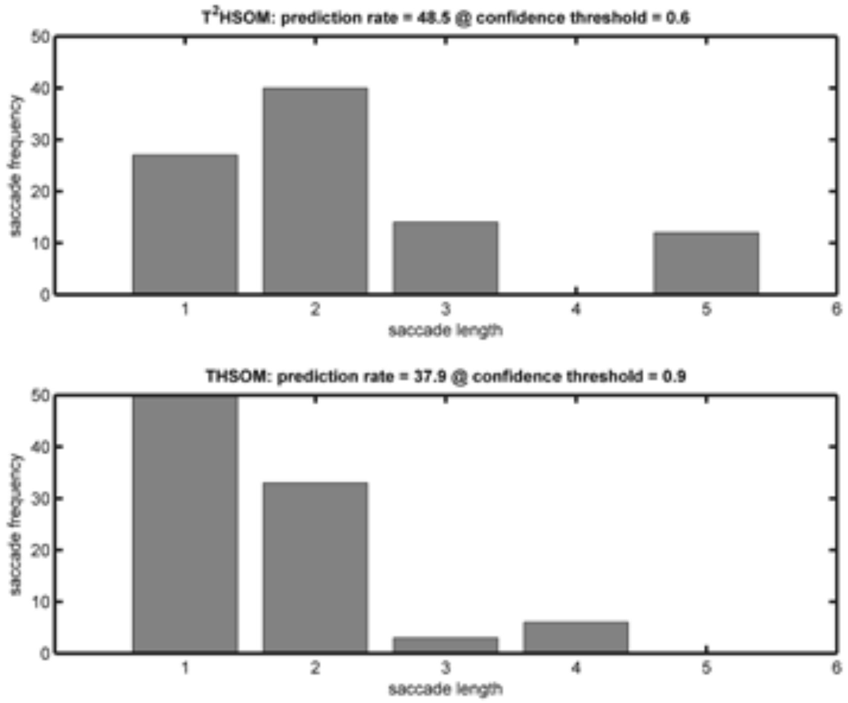


FIGURE 6. PREDICTION ACCURACY AND CHUNK LENGTH FOR DISTRIBUTED AND LOCALIST LEARNING

5.4 Simulation 3: string alignment

Figure 7 shows the temporal activation paths fired by the six present indicative forms of the Italian verb CREDERE ‘believe’ after training. Note that all paths share the sequence ‘#→C→R→E→D’ corresponding to the morphological root of the verb. Different paths start branching out upon the transition from the root to its inflectional endings, eventually giving rise to six distinct (partially overlapping) node sequences: ‘O→#’, ‘I→#’, ‘E→#’, ‘I→A→M→O→#’, ‘E→T→E→#’, ‘O→N→O→#’. Activation paths thus correspond to morphological constituents. The overall resulting structure of the map is morphologically interesting in several respects.

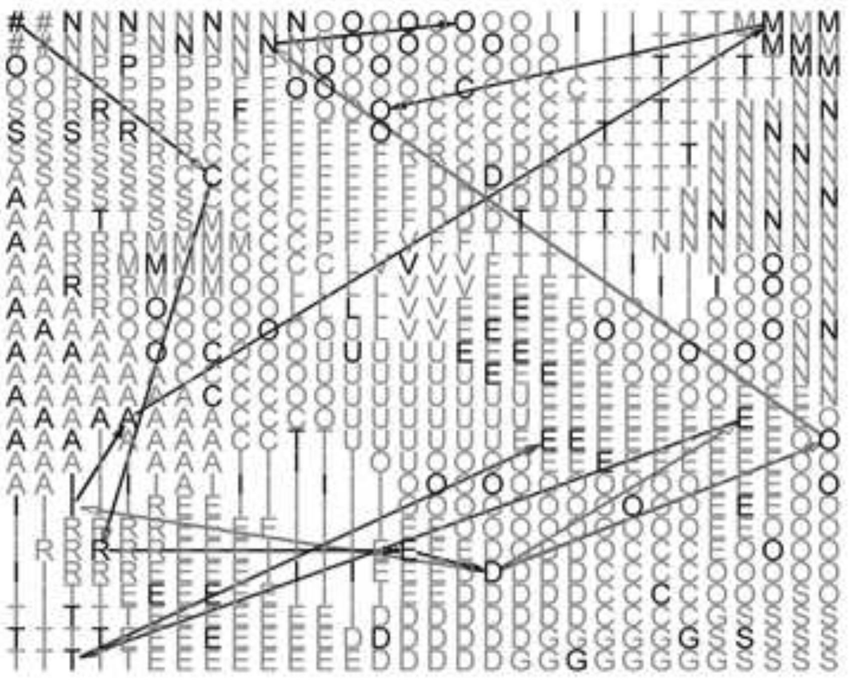


FIGURE 7. TEMPORAL ACTIVATION PATHS OF CREDERE PRES IND FORMS

First, let us look at the way paradigmatically homologous forms such as *vediamo* ‘we see’ and *crediamo* ‘we believe’ are represented as activation chains on the map (Figure 8). The two BMU chains are fairly clearly separated on the roots *cred-* and *ved-*, but tend to converge as soon as more letters are shared by the two input forms. Eventually the substring -*iamo* leaves two BMU chains that run in parallel through the map at a very short topological distance. We take this to mean that the two substrings are recognized by the map as two instances of the same type of inflectional ending. Note that the shared substring *iamo* takes different positions in the two forms, starting from the forth letter in *vediamo* and from the fifth letter in *crediamo*. In traditional positional coding, this raises an alignment problem. In our map, the substring in question receives different but converging representations, as order information is relative rather than absolute.

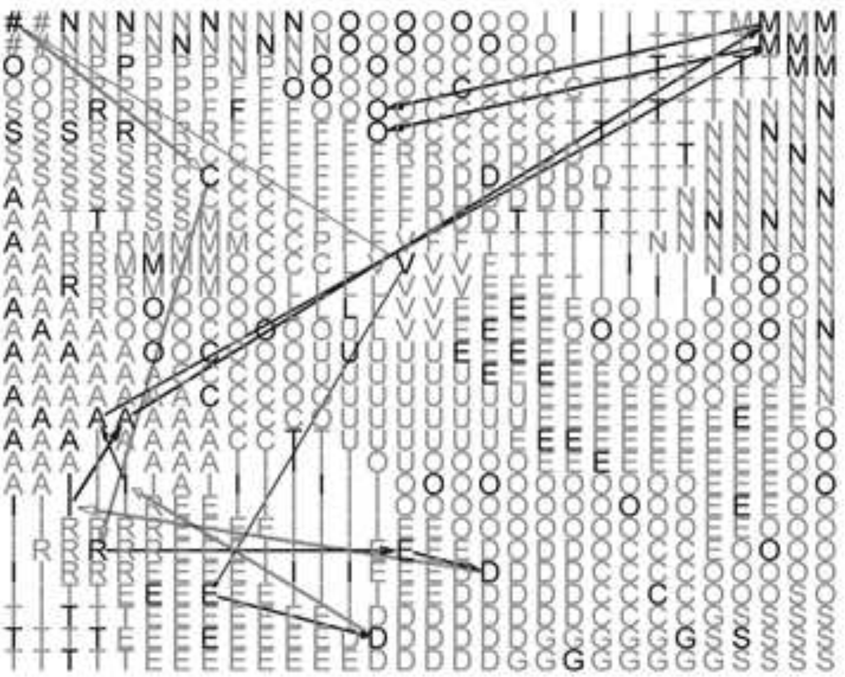


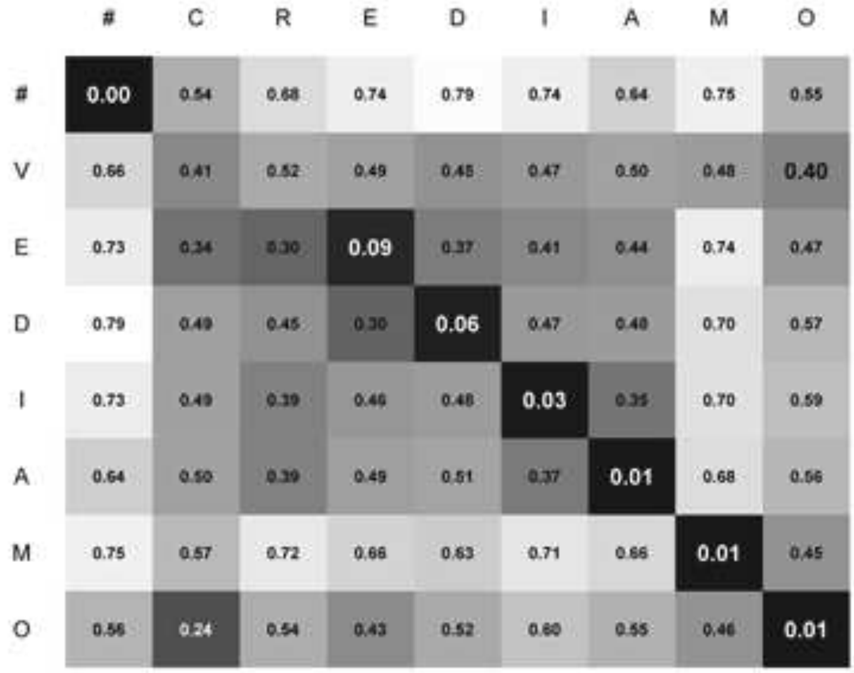
FIGURE 8. TEMPORAL ACTIVATION PATHS FOR *CREDIAMO* AND *VEDIAMO*

Analytically, convergence can be expressed in terms of topological distance between BMUs on the map. Figure 9 gives the per-node topological distance (weighted by spatial distance) of the BMU chains of *vediamo* and *chiediamo*. As the chains unfold, per-node distance progressively narrows down. In morphological terms, topological convergence expresses shared morphological structure. Note furthermore that structure is inherently "graded" at morpheme boundaries, with an early start corresponding to the shared substrings *-ed* in the roots *ved-* and *cred-*.

More difficult cases of alignment arise in the context of Semitic morphologies, where the relative position of the letters shared by morphologically-related forms can vary dramatically, as in *KaTaBa vs. yaKTuBu*, respectively the perfective and imperfective forms of the verbal triliteral root *KTB* 'write'.

An interesting question is to what extent the activation paths corresponding to Arabic perfective and imperfective forms are successful in representing the morphological notions of triconsonantal root and interdigitated vowel pattern. The problem is far from trivial, as discontinuous morphological patterns are known to be beyond the reach of chaining models for serial order. Given two perfective forms like *kataba* and *hadama*, for example, vowels in the two strings are all preceded by different left contexts.

We trained a 25x25 map on 12 perfective and imperfective transliterated Arabic forms. The two BMU chains for the forms *kataba* and *hadama* are given in Figure 10.



	#	C	R	E	D	I	A	M	O
#	0.00	0.54	0.68	0.74	0.79	0.74	0.64	0.75	0.55
V	0.66	0.41	0.52	0.49	0.45	0.47	0.50	0.48	0.40
E	0.73	0.34	0.30	0.09	0.37	0.41	0.44	0.74	0.47
D	0.79	0.49	0.45	0.30	0.06	0.47	0.48	0.70	0.57
I	0.73	0.49	0.39	0.46	0.48	0.03	0.35	0.70	0.59
A	0.64	0.50	0.39	0.49	0.51	0.37	0.01	0.68	0.56
M	0.75	0.57	0.72	0.66	0.63	0.71	0.66	0.01	0.45
O	0.56	0.24	0.54	0.43	0.52	0.60	0.55	0.46	0.01

FIGURE 9. TOPOLOGICAL DISTANCE MATRIX FOR *CREDIAMO* AND *VEDIAMO*

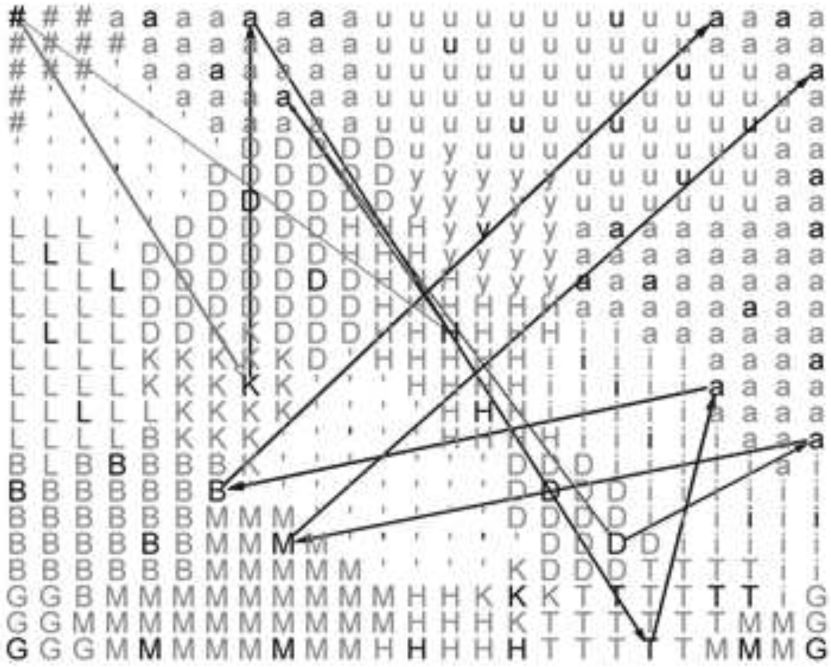
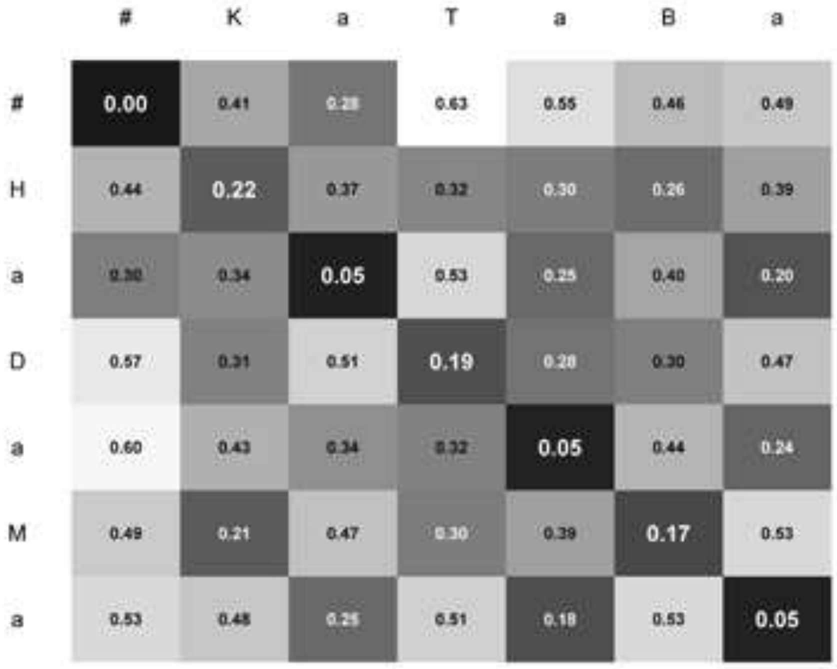


FIGURE 10. TEMPORAL ACTIVATION PATHS FOR *KATABA* AND *HADAMA*

Note that, in spite of their being preceded by different left contexts, pairs of *a* taking the same position in the two strings (e.g. *a* in second position from the left) appear to trigger topologically closer nodes on the map. The matrix of per node topological distances throws this trend in sharp relief (Figure 11). Note that *a*'s which are not time-aligned trigger nodes that are located at a considerably longer distance on the map. We shall return to a detailed analysis of this behaviour in the general discussion.



	#	K	a	T	a	B	a
#	0.00	0.41	0.28	0.63	0.55	0.46	0.49
H	0.44	0.22	0.37	0.32	0.30	0.26	0.39
a	0.30	0.34	0.05	0.53	0.25	0.40	0.20
D	0.57	0.31	0.51	0.19	0.28	0.30	0.47
a	0.60	0.43	0.34	0.32	0.05	0.44	0.24
M	0.49	0.21	0.47	0.30	0.39	0.17	0.53
a	0.53	0.48	0.25	0.51	0.18	0.53	0.05

FIGURE 11. TOPOLOGICAL DISTANCE MATRIX FOR *KATABA* AND *HADAMA*

5.5 Simulation 4: STM and LTM dynamics in THSOMs

We tested a THSOM on a task of Immediate Serial Recall (ISR) of learned strings. The THSOM trained for Simulation 2 above is shown one form at a time, randomly extracted from the same training corpus (66 Italian verb forms) and is immediately asked to recall it. The test is repeated over again, by showing the map more forms, each followed by immediate recall. The simulation is intended to model a few aspects of STM and LTM interaction. In a THSOM, LTM processes have two functions: i) node specialization to context-sensitive recognition of specific symbols, ii) consolidation of Hebbian temporal connections between consecutively activated nodes. The STM dynamic, on the other hand, consists in distributed transient activation of map nodes prompted by i) their spatial connections to input vector representations and ii) their temporal connections to other map nodes.

In the SOM literature, node activation is typically assumed to be sustained over one time step, until the next input stimulus is shown to the map. In fact, we have evidence that human subjects exposed to an arbitrary string in a classical ISR task can sustain, for a few seconds, the simultaneous activation of more symbols in the string. We thus modelled concurrent sustained activation by integrating the ST activation patterns iteratively triggered by each symbol in an input string. Upon recall, the THSOM must be able to reproduce the original input string on the basis of the (ST) integrated activation state. This is far from trivial, as pattern integration apparently has no direct order information. Figure 12 shows the map's per word recall accuracy at different learning epochs (50, 70 and 100), as a function of an increasingly filtered ST integrated activation. Filtering an integrated activation pattern means focusing on just shown symbols by getting rid of concurrent spurious node activations. The experiment gives a measure of the robustness of integrated activation in conveying information about order and items in the string to recall and of the importance of LT learning for ST recall.

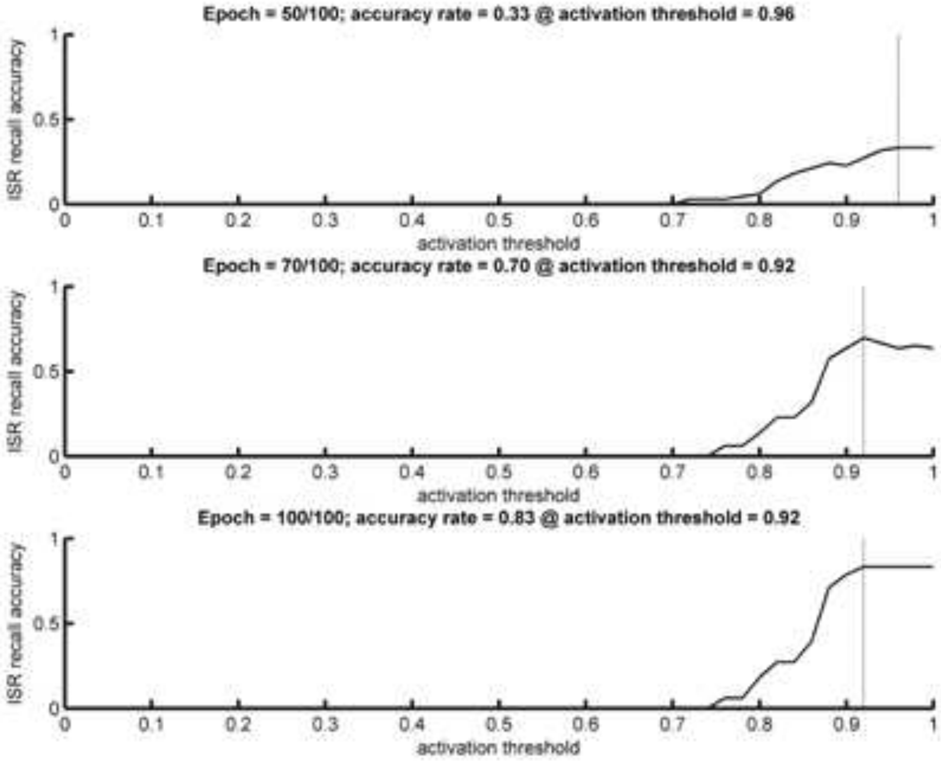


FIGURE 12. ISR ACCURACY BY INCREASINGLY THRESHOLDED ACTIVATION AT DIFFERENT LEARNING EPOCHS

6. GENERAL DISCUSSION

THSOMs define an interesting class of general-purpose dynamic memories for serial order. They learn to encode and decode variable-length sequences using minimally-entropic hierarchical structures. The structures can easily be inspected by observing temporal activation peaks (BMUs' temporal chains) in a map being exposed to a particular sequence of stimuli. In fact, more than just storage is involved here. In what follows, we discuss a few representation and processing aspects of THSOM memory mechanisms.

6.1 Conjunctive coding

A trained THSOM behaves like a variable-order stochastic Markov chain, with inter-node connections building expectations about possible upcoming strings. In a THSOM, individual symbol types are represented as distributed patterns of topologically organized nodes. The distance between any two nodes in the map space is an inverse function of two values: i) the correlation between their input connections, ii) the correlation of their pre-synaptic connections (their average left-context). In this respect, distributed activation patterns represent conjunctive codes for symbols (Simulation 1). Yet, we observe that distributed representations of this kind are not purely token-based, as all nodes dedicated to the same symbol embody an invariant representation of that symbol (the vector of spatial connection weights), with context-sensitivity being encoded through patterns of Hebbian connections. This allows for important type-based generalizations to be made, as illustrated by Simulation 3 above on word alignment (see also section 6.4 below).

6.2 Chunking

The way sequences are dynamically stored in a THSOM has an influence on its processing behaviour. THSOMs can learn to "chunk" frequently attested strings by processing them through "dedicated (unique) temporal chains" of BMUs (Simulation 1). As we saw, this is the result of a long-term memory process, based on incremental weight adjustment over Hebbian connections. The process has a significant impact on the map's on-line short-term dynamics. By maximizing the number of dedicated chains, a THSOM maximizes its "predictive ability", since non-intersecting BMU chains are memory structures that can be traversed with certainty after activation of the first node in the chunk (Simulation 2). This increases fluency and speed in on-line processing (e.g. in reading performance, as shown by Ferro *et al.*, 2010). Finally, it may also increase the capacity of the STM span, since more biased expectations make order relationships easier to retrieve from integrated short-term activation patterns (Simulation 4).

Chunking depends on how good the map is in keeping memory of past events (the so-called memory order of the map). Since words are made up by recurrent, limited combinations of letters, they tend to activate the same sequences of nodes over again, thus causing BMU chains to merge. In Simulation 1, for example, ‘ABC’ and ‘BCA’ share the substring ‘BC’. For the map to be able to devote distinct BMU chains to the two strings, it has to keep track of the different (left) contexts where ‘BC’ occurs. Chunking is thus a function of the memory order of the map. Recall that distributed temporal learning (Pirrelli *et al.*, 2010) is based on first-order Hebbian connections only. Despite this limitation, chunking extends over longer stretches of symbols due to cascaded propagation of context-sensitive activation, thereby simulating orders of memory greater than one (Simulation 1). This is achieved through a profligate use of memory resources allowing for storage of “context-sensitive redundant information”, in sharp contrast with classical lexical architectures where storage parsimony is at a premium.

6.3 Serial recall: from time to space

Another important aspect of the way THSOMs store information is that they can transcode “time into space”. Time-bound order information defines precedence relations between symbols in a sequence and plays an important role in node activation. Recall that two nodes that are sensitive to the same symbol tend to stay close in the map space. Moreover, they tend to specialize for differing occurrences of that symbol in context, depending on their Hebbian connections to past BMUs. An important implication of this point is that it becomes possible to “decode” order information from patterns of short-term node activation only. In Simulation 4 we put this behaviour to a challenging empirical test, by modelling Immediate Serial Recall as the task of retrieving item and order information from the integrated pattern of short-term activation prompted by a recently presented string of letters. Although the evidence reported here is admittedly anecdotal, the simulation is a promising attempt at modelling LTM influence on STM processes. A striking feature of the model is that long-term chunking makes room for more reliable predictions, thus making it easier for the map to elicit unambiguous item and order information from integrated short-term activation patterns. Unlike all other models we are aware of, the result is based on the non-trivial interaction between two very different mechanisms for STM and LTM: respectively, transient node activation and trained Hebbian connections.

6.4 Word alignment

Map nodes encode both spatial information, about a symbol type, and temporal information, about the symbol's context of use. Spatial information makes it possible for a node to ignite every time its mostly correlated symbol is shown, irrespective of its position. Temporal information encodes information about the relative position of a symbol in the input string. If both space and time concur in the topological organization of a THSOM, as is the case of the training regime adopted for our simulations here, conflicting requirements engage in a competition for primacy. The resulting long-term specialization of symbol-specific nodes for order-sensitive information has interesting repercussions on issues of "word alignment" (Simulation 3).

Two input strings like '#CREDIAMO' and '#CREDO' (Figure 7) share the same chain of activated units up to 'D'. This is a consequence of the fact that two letters are indistinguishable for the map if they occur in identical left contexts. Thus, in this case letter alignment corresponds to BMU sharing. What happens when left contexts are different? In '#VEDIAMO' and 'CREDIAMO' (Figure 8) the letters 'E' and 'D' are recognized by two different nodes lying close to each other on the map. Although the two resulting BMU chains develop independently, they nonetheless run in parallel, showing a tendency to converge. Admittedly, this is a weaker criterion for alignment than BMU chain sharing, but is more widely applicable across morphological families. Arabic discontinuous roots, to give but one example, do not possibly share the same chain of BMUs. 'K' in *yaKtubu* normally triggers a distinct node than 'K' in *Kataba* would do, as the two 'K' are embedded in different local contexts, prompting different expectations on possibly ensuing letters. Nonetheless, since they are instances of the same type, they will activate two relatively close nodes on the map.

A more difficult case of word alignment arises in connection with two perfective forms like *kataba* and *hadama*, where the same letter *a* occurs in three different contexts. How is it possible for the map to discriminate them? How can the map possibly align them according to their position in the two strings if they are preceded by different letters? In fact, this seems to be a necessary step to take if we want the map to get a notion of the Arabic perfective vowel pattern. To understand how this is possible, observe that temporal information is not limited to information about the actually occurring left context. The BMU activated by the symbol 'A' in the input string '#HA' at time *t* receives support, through temporal connections, from all nodes activated at time *t-1*. The nodes include, among others, the 'K' node, which competes with the 'H' node at time *t-1* as it receives temporal support from the '#' node activated at time *t-2* (due to the existence of

‘#KA’ in *kataba*). By reverberating simultaneous activation of competing nodes to an ensuing state, the map can place ‘A’ nodes triggered by ‘#KA’ and ‘#HA’ in the same area, as they share a comparatively large portion of pre-synaptic support. In general, the mechanism allows the map to keep together nodes activated by letters in the same position in the string, as shown in Figure 13 below.

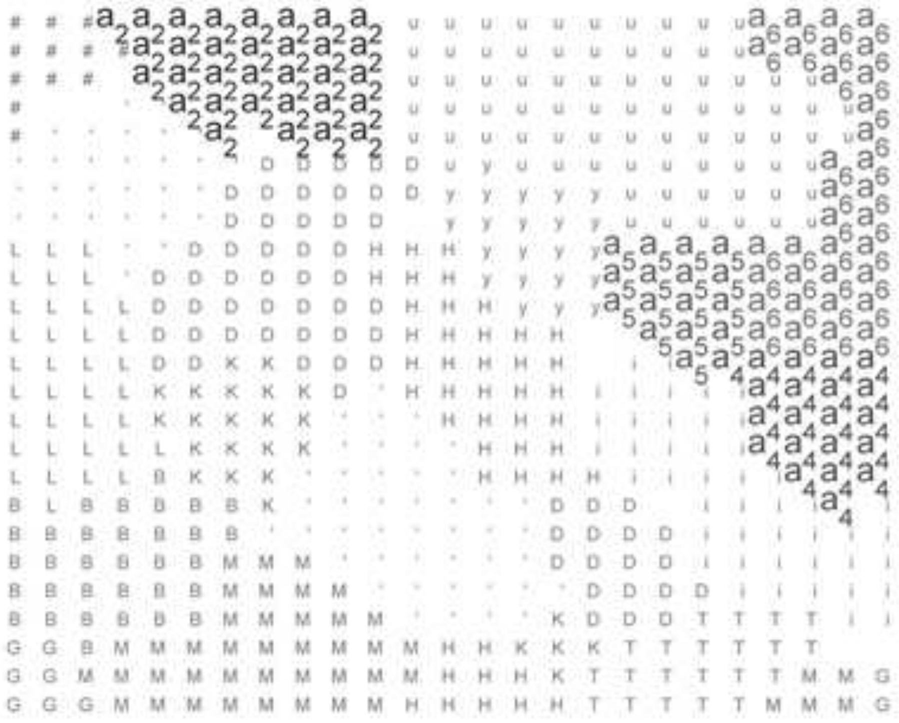


Figure 13. Clusters of *a* indexed by position in Arabic vowel patterns

6.5 Lexical hierarchies and morphological structure

THSOMs are strongly biased towards developing lexical hierarchies (see Simulation 1) that are strongly reminiscent of morphological paradigms (Pirrelli *et al.*, 2010). For example, Simulation 3 shows that all present indicative forms of the verb CREDERE activate a unique chain of BMUs corresponding to the verb’s lexical root (*cred-*). Upon hitting the final letter of the root, independent activation chains start branching out through post-synaptic transitions. As more alternative chains are competing with one another, transition probabilities reflect the relative frequency of paradigmatically-related verb forms.

Figure 14 illustrates the overall effect of such competition in the CREDERE present indicative paradigm. The vertical axis in the two panels shows where all BMUs lie on the map in terms of *x* coordinates (upper panel) and *y* coordinates (lower panel). The horizontal axis in both panels give the absolute position of each letter in the string. Solid lines represent inter-BMUs transitions, with line thickness proportional to their conditional probability. The resulting trend in entropy marks the presence of morphological structure: sub-lexical constituents (roots and endings) show a path of considerably thicker transitions than constituent boundaries. Not surprisingly, morphologically complex forms give rise to branching out structures, but their individual constituents exhibit lower entropy levels.

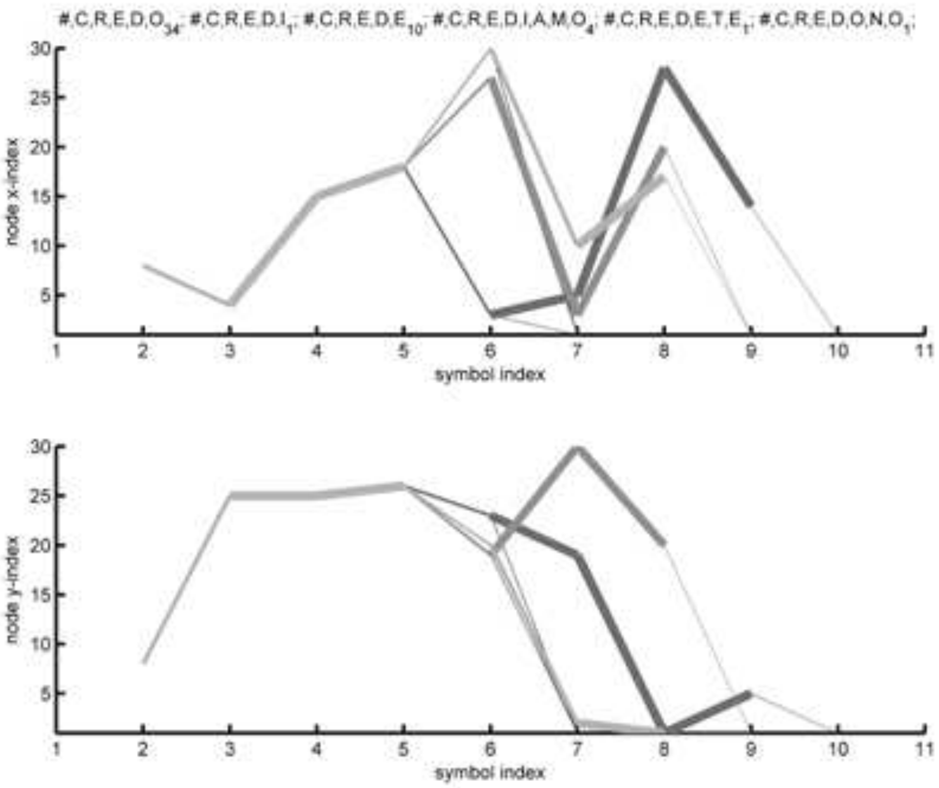


FIGURE 14. TRANSITION PROBABILITIES IN CREDERE PRES IND PARADIGM

Note that hierarchical structures of the kind depicted in Figure 14 enforce "concurrent storage" of paradigmatically-related forms, thus modelling "frequency-based competition" between base and derived forms, rather than deriving the latter from the former through on-line processing. They are thus readily amenable to being used as computer models of well-known paradigm-based effects in lexical decision tasks such as "family size" and "family frequency" effects (Moscoso del Prado Martín *et al*., 2004). Secondly, they appear to have the potential for replicating effects of automatic decompositional processing in mono-morphemic words like *brother*, due to partial overlapping with the morphologically-unrelated entry *broth*. Finally, the development of minimally-entropic strings of symbols is known to be an important determinant of linguistic structure, thus suggesting the importance of predictive behaviour in language processing and the primacy of frequency and sequentiality over hierarchical constituency in line with evidence offered by Bybee (2002), Reh (1986), Christiansen & Chater (1999) and Santelmann & Jusczyk (1998).

7. CONCLUDING REMARKS

This paper presents a computational class of lexical memories based on Kohonen's Self-Organizing Maps, augmented with a temporal layer of synaptic connections encoding order information. This class of models, dubbed THSOMs (Koutnik, 2007; Pirrelli *et al*., 2010; Ferro *et al*., 2010), addresses a number of interesting desiderata for lexical memories, including non trivial aspects of dynamic interaction between STM and LTM processes.

THSOMs can be trained on both symbol codes and their order in the input string by incrementally storing this knowledge respectively through topologically organized map nodes on the one hand and Hebbian inter-node synaptic connections on the other hand. Nodes that are sensitive to the same input symbol tend to get specialized for particular instances of the symbol in context. This is reminiscent of conjunctive coding in both classical and recurrent connectionist architectures. Unlike connectionist conjunctive representations, however, where both order and item information is collapsed on the same layer of connectivity, THSOMs keep the two sources of information stored on separate layers: the temporal layer and the spatial layer respectively. The aspect has interesting repercussions on issues of order-independent generalizations over symbol types and goes a long way to addressing both dispersion and alignment problems in word matching.

THSOMs are primarily "memory devices" and can simulate morphologically-related phenomena of lexical organization in an interesting way. It is important to appreciate in this connection that the stored content of a THSOM can be monitored "independently" of any input representation or running task. In principle, it is possible to inspect the hierarchical structure of memorized words by navigating the network of nodes and their associative connections. THSOMs are trained to memorize strings incrementally, one letter at a time. Likewise, internal memory structures can be accessed and retrieved incrementally, by scanning the input string letter by letter, from left to right. When a progressively longer portion of a word is let through, multiple BMU paths may be activated concurrently, representing "expectations" that are consistent with what is shown to the map. Surely, multiple alternatives get narrowed down as soon as more of the input string is shown to the map, in keeping with "cohort models" of lexical access (Marslen Wilson, 1990).

The recent use of classical multilayered perceptrons and recurrent variants thereof as models of lexical memory (Botvinik & Plaut, 2006; Sibley *et al*., 2008) falls short of offering a comparable battery of processing mechanisms and organizational principles. First, they are arbitrary input-output pattern associators, their hidden patterns of synaptic associations being elicited only upon presentation of an input trigger. Hence, it makes comparatively little sense to probe the internal state of a multi-layered perceptron incrementally, by showing a progressively larger portion of an input string. Furthermore, the network's behaviour is crucially determined by the whole pattern of concurrent input activation and it cannot build up expectations about possible continuations of the incoming input string. In fact, Sibley and colleagues simulate lexical storage by training a recurrent network on a task of Immediate Serial Recall (ISR). What the network appears to memorize is a set of activation patterns arbitrarily mapping input representations onto output representations. When input and output representations happen to be identical, as in an auto-encoding task, the nature of stored information does not change. The model is learning to practise ISR, rather than using the task to probe an independent memory content.

We may eventually wonder what role morphological structure plays in THSOMs. As we saw, THSOMs tend to use redundant, "context-sensitive information" to maximize the number of dedicated BMU chains. This not only maximizes the ability of a map to predict upcoming letters in the input string, but also increases the order of its STM span by reducing the number of alternative paths that are concurrently sustained. The effect is achieved in THSOMs with a simple "predictive drive". The network tends to maximize prediction accuracy. This entails maximizing discriminability in the input space, which in turn determines the formation of the maximum possible number of BMU chains (the THSOMs "chunking" mechanism), which at the same time lowers the demands of STM and the necessity to sustain several competing memory traces. In other words, in THSOMs there is a continuous interplay of memory, encoding and processing mechanisms, under the pressure of prediction accuracy.

This view prompts an interesting reappraisal of the role of morphological structure in lexical organization and processing. In THSOMs, morphological formatives (whether affixal or templatic) emerge as graded temporal chains of BMU activation. Only in some cases, shared morphological structures are directly mirrored by shared BMU chains, as in the case of root-sharing inflected forms of the same verb (*credo* and *crediamo*). In other cases, BMU chains which are fired by the same formatives are not identical. Rather, they unfold independently, running in parallel through the same map areas, as in the case of *vediamo vs. crediamo* (Figure 8). "Bundles of parallel chains" of this kind represent the closest possible correlate to the notion of morphological formative in a THSOM. Crucially, the correlate is not the result of a process of redundancy-free data compression, but rather the outcome of topological organization based on both spatial and temporal information.

Parallel chains are considerably more general morphological structures than shared chains, as they apply to discontinuous formatives (e.g. Arabic triconsonantal roots). Bundles of parallel chains thus define deeply entrenched, probabilistically-supported transitions which the map repeatedly fall into during word processing. Their role is to constrain the map's response by preventing activation chains from getting astray in the presence of noisy, rare or novel input words. For example, a THSOM that misidentifies a single input character does not catastrophically mistake all ensuing characters (as predicted by associative chaining). Rather it tends to recover from an early mistake by relying on internalized expectations based on the amount of word structure in the training data. In this sense, predictive expectations and probabilistically supported generalizations are only two sides of the same coin: optimal memory organization for efficient word recognition and production.

ACKNOWLEDGEMENTS

We gratefully acknowledge funding from the European's Community Seventh Framework Programme under grant agreement no FP7-231453 (HUMANOBS, Humanoids That Learn Socio-Communicative Skills Through Observation) and from the National Research Council of Italy, under 2010 grant agreement "Paradigmi Morfologici" for Italo-French cooperation.

REFERENCES

Aitchison, J. (1994). *Words in the Mind: An Introduction to the Mental Lexicon*. Blackwell: New York.

Alegre, M. & Gordon, P. (1999). Frequency effects and the representational status of regular inflections. *Journal of Memory and Language* 40, 41-61.

Altmann, G.T.M. & Kamide, Y. (1999). Incremental interpretation at verbs: restricting the domain of subsequent reference. *Cognition* 73, 247-264.

Baayen, H. (2007). Storage and computation in the mental lexicon. In G. Jarema & G. Libben (Eds.), *The Mental Lexicon: Core Perspectives* (pp. 81-104), Amsterdam, Elsevier.

Baayen, H., Dijkstra, T. & Schreuder, R. (1997). Singulars and plurals in Dutch: Evidence for a parallel dual route model. *Journal of Memory and Language* 37, 94-117.

Baddeley, A.D. (1964). The redundancy of letter sequences and space information.

American Journal of Psychology 77 (2), 322.

Baddeley, A.D. (1966). Short-term memory for word sequences as a function of acoustic, semantic and formal similarity. *Quarterly Journal of Experimental Psychology* 18(4), 362-365.

Baddeley, A.D. & Hitch, G. (1974). Working memory. In G.H. Bower (Ed.), *The Psychology of Learning and Motivation: Advances in Research and Theory*, Vol. 8 (pp. 47-89). New York: Academic Press.

Baddeley, A.D., Thomson, N. & Buchanan, M. (1975). Word length and the structure of short term memory. *Journal of Verbal Learning and Verbal Behavior* 14, 575-589.

Baddeley, A.D. (1986). *Working Memory*. New York: Oxford University Press.

Baddeley, A.D. & Gathercole, S.E. (1992). Learning to read: The role of the phonological loop. In J. Alegria, D. Holender, J.J. de Morais & M. Radeau (Eds.), *Analytic Approaches to Human Cognition* (pp. 153-168). Amsterdam: Elsevier.

Baddeley, A.D. (2000). The episodic buffer: a new component of working memory? *Trends in Cognitive Science* 4, 417-423.

Baddeley, A.D. (2003). Working memory: Looking back and looking forward. *Nature Reviews Neuroscience* 4 (10), 829-839.

Baddeley, A.D. (2006). Working memory: an overview. In S. Pickering (Ed.), *Working Memory and Education* (pp. 1-31). New York: Academic Press.

Baddeley, A.D. (2007). *Working Memory, Thought and Action*. Oxford: Oxford University Press.

Balota, D.A. (1994). Visual word recognition: The journey from features to meaning. In M. Gernsbacher (Ed.), *Handbook of Psycholinguistics* (pp. 303-358). San Diego: Academic Press.

Blevins, J.P. (2006). *Word-based Morphology*. *Journal of Linguistics* 42, 531-573.

Botvinick, M. & Plaut, D.C. (2006). Short-term memory for serial order: A recurrent neural network model. *Psychological Review* 113, 201-233.

Brown, G.D.A., Preece, T. & Hulme, C. (2000). Oscillator-based memory for serial order. *Psychological Review* 107, 127-181.

Burgess, N. & Hitch, G.J. (1996). A connectionist model of STM for serial order. In S.E. Gathercole (Ed.), *Models of Short-term Memory* (pp. 51-71). Hove: Lawrence Erlbaum.

Burzio, L. (2004). Paradigmatic and syntagmatic relations in italian verbal inflection. In J. Auger, J.C. Clements & B. Vance (Eds.), *Contemporary Approaches to Romance Linguistics* (pp. 17–44). Amsterdam/Philadelphia: John Benjamins.

Bybee, J.L. (2002). Cognitive processes in grammaticalization. In M. Tomasello (Ed.), *The New Psychology of Language*, Vol. II (pp. 145-167). New Jersey: Lawrence Erlbaum.

Christiansen, M.H. & Chater, N. (1999). Toward a connectionist model of recursion in human linguistic performance. *Cognitive Science* 23 (2), 157-205.

Clahsen, H. (1999). Lexical entries and rules of language: a multidisciplinary study of german inflection. *Behavioral and Brain Sciences* 22, 991-1060.

Clahsen, H. (2006). Dual-mechanism morphology. In K. Brown (Ed.), *Encyclopedia of Language and Linguistics,* Vol. 4 (pp. 1-5). Amsterdam: Elsevier.

Coltheart, M., Rastle, K., Perry, C., Langdon, R. & Ziegler, J. (2001). DRC: A Dual Route Cascaded model of visual word recognition and reading aloud. *Psychological Review* 108, 204-256.

Conrad, R. (1960). Acoustic confusion in immediate memory. *British Journal of Psychology* 51, 45-48.

Corbett, G. & Fraser, N. (1993). Network Morphology: a DATR account of Russian nominal inflection. *Journal of Linguistics* 29, 113-142.

Cowan, N. (2001). The magical number 4 in short-term memory: A reconsideration of mental storage capacity. *Behavioral and Brain Sciences* 24, 87-185.

Cowan, N. (1993). Activation, attention and short-term memory. *Memory and Cognition* 21, 162-167.

Daugherty, K. & Seidenberg, M.S. (1992). Rules or connections? The past tense revisited. In *Proceedings of the 14th Annual Meeting of the Cognitive Science Society* (pp. 259-264). Hillsdale, NJ: Lawrence Erlbaum.

Davis, C.J. & Bowers, J.S. (2004). What do Letter Migration Errors Reveal About Letter Position Coding in Visual Word Recognition? *Journal of Experimental Psychology: Human Perception and Performance* 30, 923-941.

DeLong, K.A., Urbach, T.P. & Kutas, M. (2005). Probabilistic word pre-activation during language comprehension inferred from electrical brain activity. *Nature Neuroscience* (Nature Publishing Group) 8, 1117-1121.

De Vaan, L., Schreuder, R. & Baayen, R.H. (2007). Regular morphologically complex neologisms leave detectable traces in the mental lexicon. *The Mental Lexicon* 2 (1), 1-24.

Dressler, W.U., Kilani-Schoch, M., Gagarina, N., Pestal, L. & Pöchtrager, M. (2006). On the typology of inflection class systems. *Folia Linguistica* 40, 51-74.

Ebbinghaus, H. (1964). *Memory: A Contribution to Experimental Psychology*. New York: Teachers College, Columbia University.

Eddington, D. (2002). Dissociation in Italian conjugations: a single-route account. *Brain and Language* 81, 291-302.

Ellis, N. & Schmidt, R. (1998). Rules or associations in the acquisition of morphology? the frequency by regularity interaction in human and pdp learning of morphosyntax. *Language and Cognitive Processes* 13, 307-336.

Elman, J.L. (1990). Finding Structure in Time. *Cognitive Science* 14 (2), 179-211 (DOI: doi: 10.1207/s15516709cog1402_1).

Estes, W.K. (1991). On types of item coding and sources of recall in short-term memory. In W.E. Hockley & S. Lewandowsky (Eds.), *Relating Theory and Data: In Honor of Bennet B. Murdock* (pp. 175-194). Hillsdale, NJ: Lawrence Erlbaum.

Federmeier, K.D. (2007). Thinking ahead: the role and roots of prediction in language comprehension. *Psychophysiology* 44, 491-505.

Ferro, M., Ognibene, D., Pezzulo, G. & Pirrelli, V. (2010). Reading as active sensing: a computational model of gaze planning in word recognition. *Frontiers in Neurorobotics* 4 (6), 1-16 (DOI: 10.3389/fnbot.2010.00006).

Ford, M., Marslen-Wilson, W. & Davis, M. (2003). Morphology and frequency: contrasting methodologies. In H. Baayen & R. Schreuder (Eds.), *Morphological Structure in Language Processing* (pp. 89-124). Berlin/New York: Mouton de Gruyter.

Frankish, C. (1985). Modality-specific grouping effects in short-term memory. *Journal of Memory and Language* 24, 200-209.

Frost, R., Forster, K.I. & Deutsch, A. (1997). What can we learn from the morphology of Hebrew? A masked priming investigation of morphological representation. *Journal of Experimental Psychology: Learning, Memory and Cognition* 23, 829-856.

Gathercole, S.E. & Baddeley, A.D. (1989). Evaluation of the role of phonological STM in the development of vocabulary in children: A longitudinal study. *Journal of Memory and Language* 28, 200-213.

Gathercole, S.E. & Pickering, S.J. (2001). Working memory deficits in children with special educational needs. *British Journal of Special Education* 28, 89-97.

Grossberg, S. (1978). A theory of visual coding, memory and development. In E.L.J. Leeuwenberg & H.F.J.M. Buffart (Eds.), *Formal Theories of Visual Perception* (pp. 7-26). New York: John Wiley and Sons.

Harm, M.W. & Seidenberg, M.S. (1999). Phonology, Reading Acquisition and Dyslexia: Insights from Connectionist Models. *Psychological Review* 106 (3), 491-528.

Hay, J. (2001). Lexical frequency in morphology: is everything relative? *Linguistics* 39, 1041-1111.

Hay, J.B. & Baayen, R.H. (2005). Shifting paradigms: gradient structure in morphology. *Trends in Cognitive Sciences* 9, 342-348.

Hebb, D.O. (1949). *The Organization of Behavior: A Neuropsychological Theory*. New York: Wiley.

Hebb, D.O. (1961). Distinctive features of learning in the higher animal. In J.E. Delafresnaye (Ed.), *Brain Mechanisms and Learning* (pp. 37-46). Oxford: Oxford University Press.

Henson, R.N. (1993). *Short-term Associative Memories*. Unpublished master thesis, University of Edinburgh.

Henson, R.N. (1998). Short-term memory for serial order: The start-end model. *Cognitive Psychology* 36, 73-137.

Hintzman, D.L. (1986). Schema abstraction in a multiple-trace memory model. *Psychological Review* 93, 328-338.

Houghton, G. (1990). The problem of serial order: A neural network model of sequence learning and recall. In R. Dale, C. Nellish & M. Zock (Eds.), *Current Research in Natural Language Generation* (pp. 287-318). San Diego, CA: Academic Press.

Houghton, G. & Hartley, T. (1995). Parallel models of serial behaviour: Lashley revisited. *Psyche* 2, 2-25.

Joanisse, M.F. & Seidenberg, M. (1999). Impairments in verb morphology after brain injury: a connectionist model. In *Proceedings of the National Academy of Sciences USA*, 7592–7597.

Jordan, M.I. (1986). *Serial order: A parallel distributed processing approach*. Technical Report No. 8604 ICS. La Jolla, CA: University of California at San Diego.

Kohonen, T. (2001). *Self-Organizing Maps*. Heidelberg: Springer.

Koutnik, J. (2007). Inductive Modelling of Temporal Sequences by Means of Self-organization. In *Proceeding of Internation Workshop on Inductive Modelling (IWIM 2007)*, 269-277.

Lashley, K.S. (1951). The problem of serial order in behavior. In L.A. Jefress (Ed.), *Cerebral Mechanisms in Behavior* (pp. 112-146). New York: Wiley.

Libben, G. (2006). Why studying compound processing? An overview of the issues. In G. Libben & G. Jarema (Eds.), *The Representation and Processing of Compound Words* (pp. 1-22). Oxford: Oxford University Press.

Longtin, C.M., Segui, J. & Hallé, P.A. (2003). Morphological priming without morphological relationship. *Language and Cognitive Processes* 18 (3), 313-334.

Lüdeling, A. & Jong, N. de (2002). German particle verbs and word formation. In N. Dehé, R. Jackendoff, A. McIntyre & S. Urban (Eds.), *Explorations in Verb-Particle Constructions*. Berlin/New York: Mouton der Gruyter.

MacWhinney, B. (2000). *The CHILDES Project: Tools for Analyzing Talk. Volume 2: The Database*. Hillsdale, NJ: Lawrence Erlbaum.

Maratsos, M. (2000). More overregularizations after all. *Journal of Child Language* 28, 32-54.

Marslen-Wilson, W. (1990). Activation, competition and frequency in lexical access. In G. Altmann (Ed.), *Cognitive Models of Speech Processing* (pp. 148-172). Cambridge, MA: MIT Press.

Matthews, P.H. (1991). *Morphology*. Cambridge: Cambridge University Press.

McClelland, J.L. & Rumelhart, D.E. (1981). An interactive activation model of context effects in letter perception. Part 1: An account of basic findings. *Psychological Review* 88, 375-407.

McClelland, J.L. & Patterson, K. (2002). Words or rules cannot exploit the regularity in exceptions (Reply to Pinker and Ullman). *Trends in Cognitive Science* 6, 464-465.

McQueen, J.M. & Cutler, A. (1998). Morphology and word recognition. In A. Spencer & A.M. Zwicky (Eds.), *The Handbook of Morphology* (pp. 406-427). Oxford: Blackwell.

Miller, G.A. (1956). The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological Review* 63 (2), 81-97.

Miller, G.A. & Chomsky, N. (1963). Finitary models of language users. In D. Luce, R. Bush & R. Galanter (Eds.), *Handbook of Mathematical Psychology,* Vol. 2 (pp. 419-491). New York: Wiley.

Moscoso del Prado Fermìn, M., Bertram, R., Häikiö, T., Schreuder, R. & Baayen, H. (2004). Morphological family size in a morphologically rich language: The case of Finnish compared with Dutch and Hebrew. *Journal of Experimental Psychology: Learning, Memory and Cognition* 30 (6), 1271-1278.

Page, M.P.A. & Norris, D. (1998). The primacy model: A new model of immediate serial recall. *Psychological Review* 105, 761-781.

Papagno, C., Valentine, T. & Baddeley, A. (1991). Phonological short-term memory and foreign-language learning. *Journal of Memory and Language* 30, 331-347.

Perry, C., Ziegler, J. C. & Zorzi, M. (2007). Nested incremental modeling in the development of computational theories: The CDP+ model of reading aloud. *Psychological Review* 114 (2), 273-315.

Pinker, S. & Prince, A. (1988). On language and connectionism: Analysis of a parallel distributed processing model of language acquisition. *Cognition* 29, 195-247.

Pinker, S. & Ullman, M.T. (2002). The past and future of the past tense. *Trends in Cognitive Science* 6, 456-463.

Pirrelli, V. (2000). *Paradigmi in morfologia. Un approccio interdisciplinare alla flessione verbale dell'italiano*. Pisa: Istituti Editoriali e Poligrafici Internazionali.

Pirrelli, V., Ferro, M. & Calderone, B. (2010). Learning paradigms in time and space. Computational evidence from Romance languages. In M. Goldbach, M.O. Hinzelin, M. Maiden & J.C. Smith (Eds.), *Morphological Autonomy: Perspectives from Romance Inflectional Morphology*. Oxford: Oxford University Press.

Plaut, D.C., McClelland, J.L., Seidenberg, M.S. & Patterson, K. (1996). Understanding normal and impaired word reading: Computational principles in quasi-regular domains. *Psychological Review* 103, 56-115.

Post, B., Marslen-Wilson, W., Randall, B. & Tyler, L.K. (2008). The processing of English regular inflections: Phonological cues to morphological structure. *Cognition* 109, 1-17.

Prasada, S. & Pinker, S. (1993). Generalization of regular and irregular morphological patterns. *Language and Cognitive Processes* 8, 1-56.

Rastle, K., Davis & M.H. (2004). The broth in my brothers brothel: Morpho-orthographic segmentation in visual word recognition. *Psychonomic Bulletin and Review* 11 (6), 1090-1098.

Reh, M. (1986). Where have all the prefixes gone? *Africanistische Arbeitspapiere* 5, 121-134.

Rueckl, J.G. & Raveh, M. (1999). The influence of morphological regularities on the dynamics of connectionist networks. *Brain and Language* 68, 110-117.

Rumelhart, D.E. & McClelland, J.L. (1986). On learning of past tenses of English verbs. In J.L. McClelland & D.E. Rumelhart (Eds.), *Parallel Distributed Processing,* Vol. 2 (pp. 216-271). Cambridge, MA: MIT Press.

Santelmann, L. & Jusczyk, P. (1998). Sensitivity to discontinuous dependencies in language learners: Evidence for processing limitations. *Cognition* 69, 105-134.

Seidenberg, M.S. & McClelland, J.L. (1989). A distributed, developmental model of word recognition and naming. In A. Galaburda (Ed.), *From Neurons to Reading* (pp. 255-305). Cambridge, MA: MIT Press.

Shallice, T. & Vallar, G. (1990). The impairment of auditory-verbal short-term storage. In G. Vallar & T. Shallice (Eds.), *Neuropsychological Impairments of Short-Term Memory* (pp. 11-53). Cambridge: Cambridge University Press.

Sibley, D.E., Kello, C.T., Plaut, D. & Elman, J.L. (2008). Large-scale modeling of wordform learning and representation. *Cognitive Science* 32, 741–754.

Slamecka, N.J. (1985). Ebbinghaus: Some associations. *Journal of Experimental Psychology: Learing, Memory and Cognition* 11, 414-435.

Stemberger, J.P. & Middleton C.S. (2003). Vowel dominance and morphological processing. *Language and Cognitive Processes* 18 (4), 369-404.

Tabak, W., Schreuder, R. & Baayen, R.H. (2005). Lexical statistics and lexical processing: semantic density, information complexity, sex and irregularity in Dutch. In M. Reis & S. Kepser (Eds.), *Linguistic Evidence* (pp. 529-555). Berlin/New York: Mouton de Gruyter.

Taft, M. (1979). Recognition of affixed words and the word frequency effect. *Memory and Cognition* 7, 263-272.

Ullman, M.T. (2004). Contributions of memory circuits to language: The declarative/procedural model. *Cognition* 92, 231-270.

Waters, G.F., Rochon, E. & Caplan, D. (1992). The role of high-level speech planning in rehearsal: Evidence from patients with apraxia of speech. *Journal of Memory and Language* 31, 54-73.

Whaley, C.P. (1978). Word-non word classification time. *Journal of Verbal Learning and Verbal Behaviour* 17, 143-154.

Whitney, C. (2001). How the brain encodes the order of letters in a printed word: The SERIOL model and selective literature review. *Psychonomic Bulletin and Review* 8, 221-243.

Wickelgren, W.A. (1965). Short-term memory for phonemically similar lists. *American Journal of Psychology* 78, 567-574.

Wunderlich, D. (1996). Minimalist Morphology: the role of paradigms. In G. Booij & J. van Marle (Eds.), *Yearbook of Morphology 1995* (pp. 93-114). Dordrecht: Kluwer.

Marcello Ferro
Istituto di Linguistica Computazionale “Antonio Zampolli” (ILC CNR)
Area della Ricerca CNR
Via G. Moruzzi 1, 56124 Pisa
Italy
e-mail: marcello.ferro@ilc.cnr.it

Giovanni Pezzulo
Istituto di Linguistica Computazionale “Antonio Zampolli” (ILC CNR)
Area della Ricerca CNR
Via G. Moruzzi 1, 56124 Pisa
Italy
e Istituto di Scienze e Tecnologie della Cognizione (ISTC-CNR)
Via San Martino della Battaglia 44, 00185 Rome
Italy
e-mail: giovanni.pezzulo@ilc.cnr.it

Vito Pirrelli
Istituto di Linguistica Computazionale “Antonio Zampolli” (ILC CNR)
Area della Ricerca CNR
Via G. Moruzzi 1, 56124 Pisa
Italy
e-mail: vito.pirrelli@ilc.cnr.it