LINGUE E LINGUAGGIO

Anno XI, N. 2, Dicembre 2012

TABLE OF CONTENTS

WORD ALIGNMENT AND PARADIGM INDUCTION

Claudia Marzi Marcello Ferro Vito Pirrelli

ABSTRACT: The variety of morphological processes attested in inflectional systems of average complexity calls for adaptive strategies of word alignment. Prefixation, suffixation, stem alternation and combinations thereof pose severe problems to unsupervised algorithms of morphology induction. The paper analyses morphological generalisation as a by-product of flexible memory self-organisation strategies for word recoding. Our model endorses the hypothesis that lexical forms are memorised as full units. At the same time, lexical units are paradigmatically organised. We show that the overall amount of redundant morphological structure emerging from paradigm-based self-organisation has a clear impact on generalisation. This supports the view that issues of word representation and issues of word processing are mutually implied in lexical acquisition.

KEYWORDS: morphological generalisation, morphological paradigms, self-organising memory, word coding, word processing.

1. INTRODUCTION

In developing computational models of word processing, three fundamental issues must be addressed: i) the nature of input representations, ii) the nature of output representations, iii) the formal relationship holding between i) and ii). All three aspects have a profound influence on the way a specific morphological task is modelled.

In mainstream connectionist models of verb inflection (Rumelhart & McClelland, 1986), input representations are encoded as patterns of activated nodes on the INPUT LAYER, representing the base form of a verb, possibly augmented with an indication of the morpho-syntactic features to be realised in output. Output representations are in turn defined as patterns of nodes activated on a distinct level of connectivity, the OUTPUT LAYER, encoding the expected inflected form. Finally, the relation linking the two levels is a mapping function projecting an input pattern onto the corresponding output pattern through a third level of connectivity known as HIDDEN LAYER. Inter-layer mapping can be highly non-linear for pairs of suppletive verb forms such as *go*-*went*, but the same mechanism is held to be in place for any input-output pair, irrespective of degrees of morphological regularity. More importantly for our present concerns, mapping requires no knowledge of the morphological structure of input and output representations. In connectionist models, morphological structure is in fact conceptualised as the epiphenomenal by-product of an identity mapping between invariant portions of input and output patterns. All of this is in sharp contrast with rule-based approaches (e.g. Pinker & Prince, 1988), whereby regulars are processed by structure-sensitive rules and irregulars are simply stored in the lexicon.

The connectionist and rule-based views have dominated the cognitive debate on morphological processing of the last 25 years, mainly focusing on the nature of the mapping relationship between input and output representations. In fact, both models appear to share two fundamental assumptions: a) the view that regular inflection is the outcome of a DERIVATIONAL RELATIONSHIP between a unique lexical base and an inflected form, with the former being preliminarily available, and the latter being produced/parsed on-line; b) the idea that both input and output representations are part of the training environment, not the end result of an acquisition process. In this paper we would like to address issues a) and b) on a different footing and assess the computational and theoretical implications of this reappraisal.

In questioning assumption a), we refer to the so-called WORD-AND-PARADIGM tradition (Matthews, 1991; Pirrelli, 2000; Stump, 2001; Blevins, 2006), according to which fully inflected forms are not derived from their lexical bases, but rather mutually related through possibly recursive paradigmatic structures defining entailment relations between surface forms (Bybee, 1995; Burzio, 2004). As to assumption b), we contend that word representations are never NEUTRAL with respect to morphological operations. Knowledge of how input and output representations are mutually related cannot be decoupled from knowledge of how input and output representations are encoded in the first place. Saying that, in German, *machen* and *gemacht* are formally related is tantamount to acknowledging that the two forms can be aligned in spite of -*mach*- occurring at different time positions in the two strings. This is a considerably different perspective on morphological processing than rule-based models are ready to take. According to the view we will entertain here, word representations are not GIVEN symbolic objects, manipulated by INDEPENDENTLY-DEFINED morphological operations like prefixation, suffixation or stem alternation. Rather, they are DYNAMICALLY RECODED TIME-SERIES, whose perception can vary depending on the context and on recurrent patterns of underlying morphological structure.

In the ensuing section, we look at how alignment issues have been dealt with in two fast-developing lines of scholarly research that have so far made comparatively sparse contact: the Machine Learning literature on (un)supervised morphology induction and the psycho-cognitive literature on serial cognition.

2. ALIGNMENT ISSUES

2.1 *Machine learning and morphology induction*

Morphology induction can be defined as the task of singling out morphological formatives within morphologically complex word forms. To linguists, the task is reminiscent of Zelig Harris' empiricist goal of developing linguistic analyses (and ultimately a linguistic ontology of word-based categories and constituents) on the basis of purely formal, algorithmic manipulations of raw training data: the so called "discovery procedures" (Harris, 1951).

Absence of categorical information (e.g. morpho-syntactic or lexical information) in the training data qualifies the discovery algorithm as UNSUPERVISED. A different conceptualization of morphological induction sees the task as a CLASSIFICATION problem. The machine learner is trained on a set of forms whose classification or mutual relation is already known, and is tested upon the ability to assign the correct class (or the appropriate mapping relationship) to word forms that were not part of the training set. In this case, the learning regime is said to be SUPERVISED.

2.1.1 Supervised models

Stochastic classification algorithms, like those based on the MAXIMUM ENTROPY PRINCIPLE (Berger, Della Pietra & Della Pietra, 1996; Ratnaparkhi, 1998), deal with the problem of finding a morph m (e.g. the root *walk*) in a word w (e.g. *walked*) as the task of estimating the probability of having the category m assigned to a constituent in w, given a representation of m in terms of a number of linguistic features: e.g., position of the candidate morph, length of the constituent, existence in a dictionary, grammatical category of adjacent constituents, etc. (Uchimoto, Sekine & Isahara, 2001). A Maximum Entropy classifier solves the problem by calculating, for each category m, the conditional probability $p(m|w)$ of having m in w, given i) an appropriate feature-based recoding of w in the training data and ii) the (maximised) overall entropy of the resulting probability distribution.[1]

[1] This means that the probabilistic model should NOT be more biased than required by the input evidence. In other words, its distance from an equiprobable distribution should be made as small as necessary for the model to predict all attested data.

Unlike Maximum Entropy models of morphological classification, MEMORY-BASED approaches to morphology induction (Daelemans & van den Bosch, 2005) classify novel input by analogy to stored exemplars. A memory-based learner assumes morphological processing to be a function of either lexical retrieval or similarity-based reasoning on representations of word forms. Computation of similarity is defined on the basis of phonological, orthographical or semantic features. To ensure that only features associated with ALIGNED symbols (letters or phonological segments) are matched, exemplar representations must be aligned preliminarily. Keuleers & Daelemans (2007) enforce alignment by associating exemplars with a syllabic template consisting of three slots: onset, nucleus and coda. Finally, the similarity between two strings *X* and *Y* is measured according to the following weighted overlapping distance:

(1) $$\Delta(X,Y) = \sum_{i=1}^{n} w_i \cdot \delta(x_i, y_i)$$

where x_i and y_i indicate the values on the i^{th} matching feature F_i taken by *X* and *Y* respectively, $\delta(x_i, y_i)$ measures the distance between the two values and w_i is the weight associated with F_i, i.e. a quantitative measure of how important that feature F_i is in assigning a class to an exemplar.

Other word similarity functions have been proposed in the literature. According to the notion of PROPORTIONAL ANALOGY (Pirrelli & Yvon, 1999), word similarity is a relation among four lexical exemplars:

(2) *steal:stealer = cheat:cheater*

with *steal-stealer* and *cheat-cheater* being pairs of lexically-related words, and *steal-cheat*, *stealer-cheater* representing pairs of morphologically-related words. More formally, Pirelli & Yvon (1999) define an analogical proportion among strings of symbols in terms of identity over (pairs of) sub-strings, as follows:

(3) $$(a_1 = u \cdot v) \wedge (a_2 = u \cdot w) \wedge (a_3 = t \cdot v) \wedge (a_4 = t \cdot w)$$

where ‘$u{\cdot}v$’ means “u concatenated with v”. Accordingly, in the proportion above, the following equations hold: u = *steal*, w = *er*, t = *cheat*, v = ε (where ε represents the empty string). To avoid compilation of morphologically spurious proportions such as *cheat:corn = cheater:corner*, each member in an analogical proportion must be a LINGUISTIC SIGN, i.e. a form-meaning lexical pair. Proportionality is assumed to hold on BOTH levels of representation simultaneously. In the case at hand, since the semantic representation of *cheater* contains an agentive marker -ER and *corner* does not, the two words will never be part of the same analogical proportion.

Albright and Hayes (Albright, 2002; Albright & Hayes, 2002) address morphological generalisation by applying the MINIMAL GENERALISATION algorithm (Pinker & Prince, 1988; Albright & Hayes, 2002) to the acquisition of inflectional patterns in Italian conjugation. The algorithm consists in collecting lexical entailments between pairs of inflected forms that stand in a specific morphological relation: e.g. PRESENT → PAST or 1-SING → INFINITIVE. For example, the two Italian forms *bado* ('I take care') and *badare* ('to take care') stand in a PRES-IND, 1-SING → INFINITIVE relation. Entailments are acquired by aligning two paradigmatically related forms such as *bado* and *badare* to the left, for their shared stem to be maximised and the remaining change to be factored out. Given any two such entailments, the goal is to extract from them a maximally specific context-sensitive rule, mapping one class of forms into the other class. The authors show that minimal rules of this kind apply accurately. Moreover, their reliability score (based on the number of forms for which the mapping rule makes the right prediction) correlates with human subjects' acceptability judgement on nonce-forms.

Connectionist models do not explicitly code morphological structure into input representations. Nonetheless, they need to recode input forms to allow recurrent morphological formatives to activate overlapping units (nodes) on both input and output layers. This is necessary for overlapping input units to eventually result in similar outputs. For example, Plunkett & Juola (1999) represent a (monosyllabic) input word through a (right-justified) CCCVVVCCC template: e.g. the word *cat* (/kAt/) is represented by the training pattern ##k##A##t, the word *ox* (/Aks/) by #####A#ks, where '#' represents an absent sound. Output words, on the other hand, are assigned the same input template augmented with two extra slots (VC) for encoding inflectional endings. Accordingly, the plural *cats* (/kAts/) is represented by ##k##A##t#s and *oxen* (/AksEn/) by #####A#ksEn. This enforces a morphologically-motivated alignment on input and output representations respectively, to the effect that a specific plural suffix is always associated with the same pattern of nodes on the output layer.

2.1.2 Unsupervised models

The task of inducing morphological knowledge from raw data (i.e. neither segmented nor classified surface strings) is operationalized as consisting of two subtasks: i) finding structure in word forms, and ii) grouping word forms on the basis of the amount of shared structure. Approaches chiefly differ in the order in which the two steps are taken.

Most methods, from seminal work (Harris, 1955; Hafer & Weiss, 1974) to more recent adaptations (Juola, Hall & Boggs, 1994; Golcher, 2006; Hammarström, 2009), carry out a (possibly preliminary) segmentation step, whereby word forms are split into candidate sublexical constituents. Such candidates are thereafter subjected to validation based on frequency distributions in a reference corpus. For example, Goldsmith's (2001) algorithm for morpheme splitting goes as follows: "take all possible splits into a stem and an affix starting from the right hedge of the word and assuming a maximum affix length of six letters". Local splits are then evaluated globally, on the basis of their overall distribution: the first 100 top ranked suffixes are chosen. After segmentation, Goldsmith sets himself the task of developing a morphological grammar defined as a set of "signatures", i.e. lists of affixes that are selectively combined with sets of stems. For example, in Goldsmith's notation, the list *Null.er.ing.s* is a signature for stems like *count*, *drink, mail* and *sing*. Signatures are reminiscent of the traditional notion of morphological paradigm and define ways of grouping verbs on the basis of the affixes they share. The partitioning of affixes into signatures is validated through a principle, Minimal Description Length, which provides a way to mathematically find an optimal trade-off between two descriptively undesirable extremes: i) a "photographic" but verbose model of the data, where each word form belongs to a signature of its own and is generated according to the probability expressed by the form's relative frequency in a reference corpus; and ii) a very short but liberal model, with one overall signature, where any verb can combine with any marker according to the product of their independent probability distributions, thus generating many word forms that are not attested (including **goed* for *went*, **stricked* for *struck*, **bes* for *is* etc.).

As an alternative to Goldsmith's approach, some scholars have tried to group word forms first, on the basis of several clustering criteria such as string edit distance (Gaussier, 1999; Yarowsky & Wicentowski, 2000; Schone & Jurafsky, 2001; Baroni, Matiasek, & Trost, 2002) or Latent Semantic Analysis (e.g. Schone & Jurafsky, 2000; Baroni, Matiasek, & Trost, 2002; Freitag, 2005), to then look for structure both within and among groups. Although preliminary clustering may considerably constrain the search space for what is common among several groups, abstracting morphological processes given a family of groups is a thorny issue, because of the number of groups and because of the number of potential morphological processes. To further constrain the search space and address the specific needs of non-concatenative or templatic morphologies, some approaches first classify graphemes into vowels and consonants (Rodrigues & Ćavar, 2005, 2007; Xanthos, 2007). Each word is then separated into a consonant skeleton and a vowel pattern, for them to be eventually assessed in the light of their frequency distributions in a corpus.

2.2 *Morphology induction and serial cognition*

The divide between supervised and unsupervised models of morphology induction somewhat mirrors the interplay between structured REPRESENTATIONS and structurally-defined OPERATIONS. Supervised algorithms lay more emphasis on representations, which are assumed to be available in training. Conversely, in an unsupervised mode, machine learning must constrain mapping operations considerably. The more knowledge-rich the available representations are, the less complex the needed operations. If machine learners have no access to the internal morphological structure of their training data, then discovery procedures must be constrained enough to be able to tell relevant data from irrelevant data.

Machine learning algorithms tend to make rather specific assumptions on either word representations (in a supervised learning mode) or string processing strategies (in an unsupervised learning mode). Indeed, for most European languages, fixed-length vector representations can be developed by aligning input words to the right. Since inflection in these languages typically involves suffixation and sensitivity to morpheme boundaries, this is a sensible step to take. However, this type of encoding presupposes considerable knowledge of the morphology in the target language and does not possibly work with prefixation, circumfixation and non-concatenative morphological processes in general. Likewise, most current unsupervised algorithms (see Hammarström & Borin, 2011 for a recent survey) model morphology learning as a segmentation task, assuming a hard-wired linear correspondence between sub-lexical strings and morphological structure. However, both highly-fusional and non-concatenative morphologies hardly lend themselves to being segmented into linearly concatenated morphemes.

Many of these issues have been raised in the psycho-cognitive literature on serial cognition. A fundamental characteristic of the human language faculty is the ability to retain sequences of symbolic items (e.g. sounds, syllables, morphemes or words), to access them in recognition and production, and to find similarities and differences among them. A key issue that must be addressed to account for such a characteristic is how speakers code for item positions. Without position coding, it is not possible to retain/recognise a simple word as *pop*, where the same letter type *p* appears to be realised as two tokens embedded in different temporal contexts, or to distinguish between two anagrams such as *cat* and *act*.

Some of the earliest psychological accounts of serial order assume that item sequences are represented as temporal 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 stimulus only. Criticism of chaining models goes back to pioneering work by Lashley in the 50's (Lashley, 1951; Houghton & Hartley, 1995; for a review). For example, in order to represent a word like '#EVERY#' as a sequence of associative links between character types, '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'. So-called CONJUNCTIVE CODING addresses the problem by anchoring a symbol to its left context, thus using distinct representations (*e.g.* the bigrams '#E' and 'VE') as instances of the same 'E' type. However, conjunctive coding 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 (Davis & Bowers, 2004), thus hindering recognition of both shared and different sequences between morphologically-related forms.

We have no room here to discuss strengths and weaknesses of models of encoding time series of symbols, as proposed in the vast literature on IMMEDIATE SERIAL RECALL and VISUAL WORD RECOGNITION (see Henson, 1998; Davis, 2010; for recent reviews). It is important to emphasise at this juncture that when we cast the problem of lexical recognition/production in terms of accessing a mental representation for a familiar word (i.e. a memory trace), the focus of investigation is shifted from the issue of PROCESSING an existing representation through rule-defined algorithmic operations to the more fundamental issue of CODING and STORING word representations in the first place. Few computational models have tried to address these issues on a principled basis (Pollack, 1990; Botvinick & Plaut 2006; Sibley *et al*., 2008; among them). Temporal self-organising lexical maps (Ferro, Marzi & Pirrelli, 2011; Pirrelli, Ferro & Calderone, 2011) try to establish a potentially useful connection between issues of serial cognition and morphology induction.

3. LEXICAL MAPS

The lexicon is the store of words in long-term memory. From this perspective, modelling lexical competence requires that preliminary issues of coding/storage of time-series of symbols are taken into account before more elaborate lexical functions, such as organisation, access and recall can possibly be addressed. Temporal SOMs (hereafter TSOMs, Ferro, Marzi & Pirrelli, 2011; Pirrelli, Ferro & Calderone, 2011) offer a promising computational framework for modelling all these issues at a considerable level of detail.

3.1 *The architecture*

The architecture in Figure 1 implements lexical encoding/storage through cascading topological maps with re-entrant temporal connections. In illustrating the architecture, it is useful to distinguish an INPUT LAYER representing the most peripheral level of input encoding, from TSOMs proper, where time relations between symbols are sampled and recoded through topological relations.

For our present concerns, individual input stimuli are letters or phonological segments arranged within words as time-bound signals. Each such unit is sampled and encoded on the input layer at discrete time intervals as a vector *X(t)* of binary values. At time *t*, *X(t)* is transferred to a *T MAP* through one-to-many, trainable connections, whose $w_{i,j}$ weights say how well the input signal is transmitted from the x_j component of the input vector to the i^{th} node on the *T MAP*.

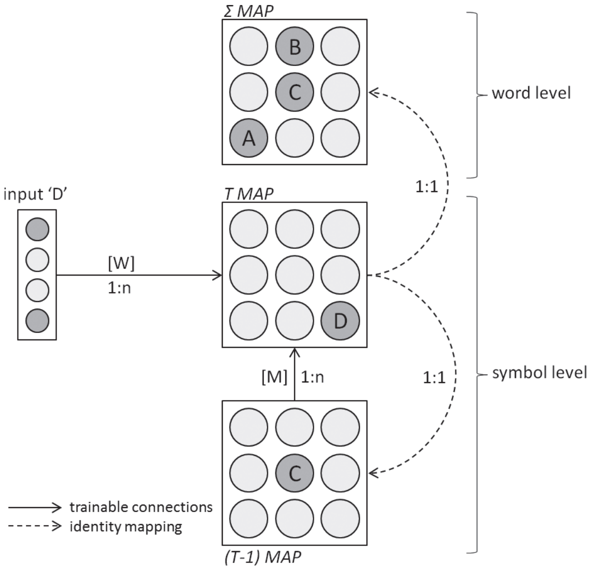


FIGURE 1. A TSOM-BASED ARCHITECTURE FOR LEXICAL ENCODING/STORAGE. A T MAP SAMPLES AND RECODES TIME-BOUND SYMBOL REPRESENTATIONS, TO INTEGRATE THEM INTO A Σ MAP, OVER WORD-LENGTH TIME INTERVALS. A (T-1) MAP, SAMPLING THE T MAP'S OUTPUT AT THE IMMEDIATELY PRECEDING TIME TICK, FEEDS BACK THE T MAP THROUGH RE-ENTRANT 'WHEN' CONNECTIONS.

Connections coming from the input layer are referred to as 'WHAT' connections, since they provide information about WHAT is shown to the map at time *t*. From a *T MAP*, the activation pattern is copied to a second map (the *(T-1) MAP* in Figure 1) through a one-to-one identical mapping function (with connection weights set to unity). The pattern is fed back to the *T MAP* at the ensuing time tick, through many-to-one, trainable connections hereafter referred to as 'WHEN' connections. The weights $m_{i,j}$ on these connections provide information on WHEN a stimulus is shown to the map. All in all, activation patterns on *T MAP* and *(T-1) MAP* define the symbol-level recoding of an input sequence. Finally, the activation pattern on a *T MAP* is copied onto a Σ MAP, which samples the input signal over word-length time intervals, thereby integrating all symbol-level patterns activated by an input word. An integrated activation pattern on the Σ MAP eventually defines word-level input recoding and represents the memory trace left by an input word.

3.2 *Recoding*

When an individual stimulus is input at time *t*, each component $x_j(t)$ on the input layer is set to either 1 or 0. All *T MAP*'s nodes are then activated concurrently as a function of WHAT is shown to the map and WHEN it is shown. The overall level of activation *H(t)* of the map at time *t* is the result of a weighted summation of the activation level $H_S(t)$, flowing through WHAT CONNECTIONS, and the contribution $H_T(t)$ of the map's expectation conveyed by WHEN CONNECTIONS:

(4) $$H(t) = \alpha \cdot H_S(t) + \beta \cdot H_T(t),$$

with *a* and *b* measuring the comparative contribution of WHAT and WHEN connections respectively, and $H_S(t)$ and $H_T(t)$ being defined as follows:

(5) $$H_S(t) = \sqrt{D} - \|1_{N\times 1} \cdot X(t) - W(t)\|,$$

(6) $$H_T(t) = M(t) \cdot H(t-1),$$

where *D* is the dimension of the input vector, *N* the size, in number of nodes, of the *T MAP*, *W(t)* the matrix of weights on WHAT connections, and *M(t)* the matrix of weights on WHEN connections.

Equation (5) says that the closer the matrix *W(t)* to the current input vector *X(t)*, the higher the $H_S(t)$ contribution to the map's overall activation level. Similarly, equation (6) says that the closer the matrix *M(t)* to the activation level *H(t-1)* of the *(T-1) MAP*, the higher the $H_T(t)$ contribution to the map's overall activation level.

Given the state of activation of a *T MAP*, the BEST MATCHING UNIT at time *t* (or *BMU(t)*) is the node with the highest activation value according to equation (4), namely:

(7) $$\hat{h}_{BMU}(t) = max_{i=1,\dots,N}\{h_i(t)\}.$$

For our present concerns, *BMU(t)* represents the output of the map at time *t*. In assessing the behaviour of a map, we compare the output to the correct response. For example, a *T MAP* is said to recode an input symbol correctly iff:

(8) $$\sqrt{\sum_{j=1}^{D}[x_j(t) - w_{BMU,j}(t)]^2 < \theta,}$$

with $\theta = 0.05$.

3.3 *Training*

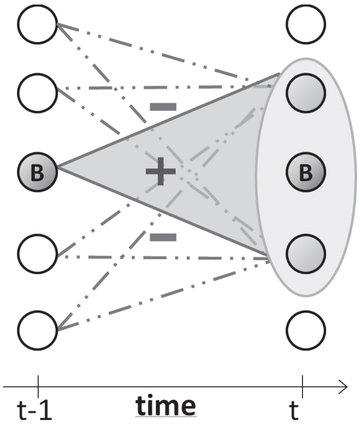


FIGURE 2. SPREADING ACTIVATION OF LONG-TERM POTENTIATION (SOLID LINES) AND LONG-TERM DEPRESSION (DOTTED LINES) OF RE-ENTRANT WHEN CONNECTIONS OVER TWO SUCCESSIVE TIME STEPS. B-NODES DENOTE BMUS.

Through repeated exposure to input stimuli, WHAT connections are attuned to specific vectors on the input layer, and WHERE connections are attuned to *(T-1) MAP*'s expectations. This is done by maximising the values associated with equations (5) and (6) above during training: WHAT connections and WHEN connections are made incrementally closer to their target values. This means that, during training, nodes become more and more sensitive to particular stimuli or classes of stimuli in specific temporal contexts. Moreover, nodes which are sensitive to similar stimuli tend to cluster together on the map. This is achieved through a propagation function which spreads connection weights from *BMU(t)* to surrounding nodes as a function of the topological distance between *BMU(t)* and the surrounding node (Figure 2), and of the current learning rate of the *T MAP*. The spreading mechanism of WHEN connections plays an important role in the generalisation bias of a TSOM. We shall return to this important point in section 5.

3.4 *Recall*

Lexical recall consists in "reading" an input word *K* off its integrated activation pattern on the Σ MAP. Intuitively, this is possible since the integrated pattern contains detailed information about a) the letters making up *K*, and b) their position in *K*.

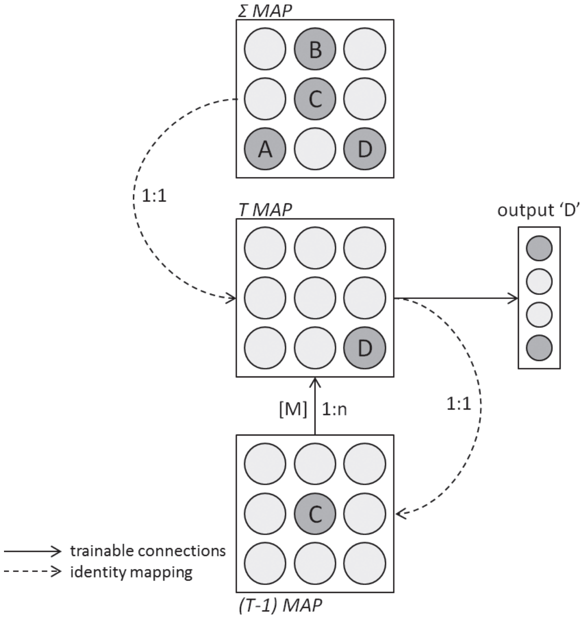


FIGURE 3. LEXICAL RECALL IN A LEXICAL TSOM ARCHITECTURE. SINGLE CHARACTERS ARE READ OFF AN INTEGRATED PATTERN ON THE Σ MAP, BASED ON *(T-1) MAP*'S RE-ENTRANT EXPECTATIONS

For any word *K* of length n_K we define its integrated activation pattern on the Σ MAP as the union set of all activation patterns triggered by the word's letters on the *T MAP*:

(9) $\widehat{H}_K = \{\hat{h}_1, \dots, \hat{h}_N\}$

and

(10) $\hat{h}_i = max_{t=2,\dots,n_K}\{h_i(t)\}\,, \qquad i = 1, \dots, N.$

Lexical recall is thus modelled using the activation function of equation (4) above, with:

$$H_S(t) = \begin{cases} \sqrt{D} - \|1_{N\times 1} \cdot X(t) - W(t)\| & , t = 1 \\ \hat{H}_K & , t = 2, \dots, n_K. \end{cases} \tag{11}$$

Recall accuracy can be measured by equation (8) above. If *K* is the input word associated with $\hat{H}_K$, then *X(t)* in (8) is the input vector encoding the t^{th} letter in *K*. This ensures that a *BMU(t)* is recalled accurately only if it matches both the correct letter and its time-stamped position in the input string.

3.5 *Alignment and generalisation*

Lexical maps can be shown to generalise to novel words. To understand how this works, it is useful to look at ACTIVATION CHAINS of morphologically-related words on a Σ MAP. The activation chain relative to the word *K* consists of all *BMU*s discharging in association with the letters in *K*. Figure 4 (left) shows *BMU* chains associated with the Italian verb forms VEDIAMO 'we see', VEDETE 'you see' (second person plural), and CREDIAMO 'we believe', with the ending -IAMO ('1st person plural present indicative') shared by two of the three forms, activating the same *BMU*s.

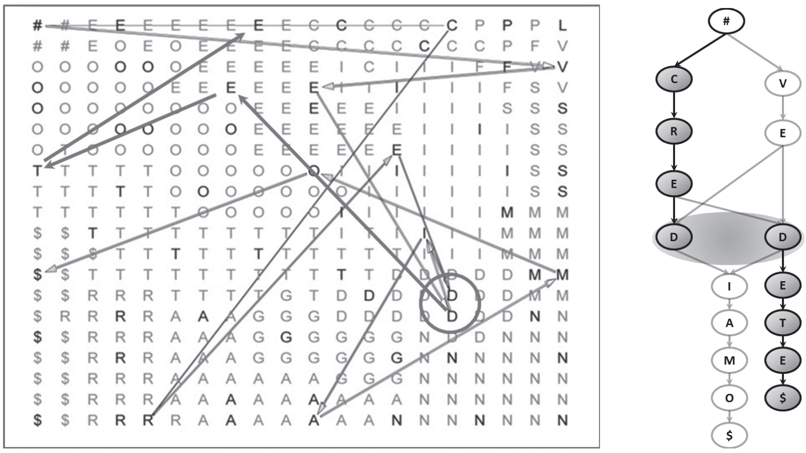
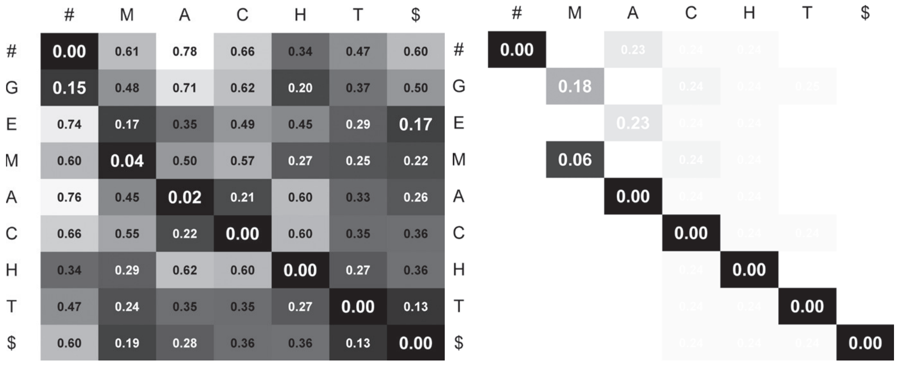


FIGURE 4. BMU ACTIVATION CHAINS FOR *VEDIAMO*-*VEDETE*-*CREDIAMO* ON A 20×20 MAP (LEFT) AND THEIR WORD-GRAPH REPRESENTATION (RIGHT).

Activation of the same *BMU*s by different input words reflects the extent to which the map perceives these words as similar. It tells us how well input words are aligned in the recoding space defined by a TSOM. The grid in Figure 5 (right) measures the difference in activation levels between pairs of *BMU*s responding to different input words. In particular, for each cell $c_{i,j}$

(12) $$c_{i,j} = \left|\hat{h}_{BMU}(i) - \hat{h}_{BMU}(j)\right|$$



	#	M	A	C	H	T	$
#	0.00	0.61	0.78	0.66	0.34	0.47	0.60
G	0.15	0.48	0.71	0.62	0.20	0.37	0.50
E	0.74	0.17	0.35	0.49	0.45	0.29	0.17
M	0.60	0.04	0.50	0.57	0.27	0.25	0.22
A	0.76	0.45	0.02	0.21	0.60	0.33	0.26
C	0.66	0.55	0.22	0.00	0.60	0.35	0.36
H	0.34	0.29	0.62	0.60	0.00	0.27	0.36
T	0.47	0.24	0.35	0.35	0.27	0.00	0.13
$	0.60	0.19	0.28	0.36	0.36	0.13	0.00

	#	M	A	C	H	T	$
#	0.00		0.23	0.24	0.24		
G		0.18		0.24	0.24	0.25	
E			0.23	0.24	0.24		
M		0.06		0.24	0.24		
A			0.00	0.24	0.24		
C				0.00	0.24	0.24	
H				0.24	0.00		
T				0.24	0.24	0.00	
$				0.24	0.24	0.24	0.00

Figure 5. topological distance matrix (left) and activation spread matrix (right) for German gemacht and macht. darker boxes signify that the corresponding letters are closer on the topological space of the map (left) and are closer in levels of activation (right). Zeros indicate that the two letters activate precisely the same node

Zero values in the grid indicate that letters in the i^{th} row and the j^{th} column of the grid activate the same node on the map. Node co-activation correlates with inter-node topological distance in the map space, as shown by the distance matrix to the left of Figure 5.

Co-activation is key to generalisation. This is illustrated in Figure 4 (right) where *BMU* chains are unfolded and arranged vertically in a WORD GRAPH. In the graph, circles are map nodes, and directed arcs represent re-entrant WHEN connections. Grey circles are nodes that are activated in association with the form CREDETE ('you believe', second person plural), under the assumption that CREDETE was not shown during training. Note that there is no direct WHEN connection from the 'D' node associated with CRED- to the *BMU* chain for the suffix -ETE ('second person plural present indicative'). Nonetheless, the latter chain is activated indirectly, due to co-activation of the the 'D' node associated with the root VED-. This shows that novel chains of *BMU*s can develop "on-the-fly" during recall as a result of levels of parallel activation spreading from time-aligned BMU chains. This is the by-product of the propagation function spreading WHEN connections from *BMU(t-1)* to *BMU(t)* and its neighbouring nodes during learning (Figure

2 above). Spreading activation of WHEN connections thus enforces the map's propensity to accept novel words by extending learned connections to local topological neighbourhoods. It should be noted, incidentally, that the entire lexicon stored in a TSOM can be represented as a huge word graph with the symbol '#' on the top node.

In previous work (Marzi, Ferro & Pirrelli, 2012b), we assessed how well this generalisation strategy works on Italian and German data by comparing two different settings of *a* and *b* in equation (4). We reported a statistically significant advantage in correctly recalling novel forms on both Italian and German, for *a = 0.5* and *b = 1*. The advantage can be explained in terms of a difference in word activation and recoding. For *a* = 0.5, the identity of the symbol shown to the map carries more weight on node activation than the symbol's timing does. In TSOMs, nodes that present comparable activation levels in association with the same stimuli are clustered in topologically connected areas. This means that neighbouring nodes will tend to be sensitive to symbol identity. Sub-clusters of nodes will be selectively sensitive to context-specific instances of the same symbol.

Sensitivity to symbol identity (as opposed to symbol timing) makes TSOMs more able to capture morphological structure. This is particularly true of concatenative morphologies, where the notion of morphological constituent is position-independent and requires the capacity to single out recurrent substrings at different positions in time. Nonetheless, highly-inflecting languages exhibit morphological processes which are not strictly concatenative. German conjugation, for example, includes suffixation (*mache/machst* 'I find/you find'), stem alternation (*finden/fanden* 'to find'/'they found'), circumfixation (*machen/gemacht* 'to make'/'made' past participle) and combinations thereof (*finden*/*gefunden* 'to find'/'found' past participle). It is thus interesting to investigate to what extent local topological propagation of WHEN connections can deal with this range of phenomena in generalisation. How can a lexical self-organising map simulate the different inductive mechanisms that are needed to deal with the entire range of morphological processes of German conjugation? Can we conceive of better generalisation strategies than local propagation? What do we understand of morphology storage and acquisition by investigating these mechanisms?

Computational simulations are helpful in addressing all these issues on a firm, empirical basis. The ensuing experiment was designed for this purpose.

4. THE EXPERIMENT

4.1 *Materials and method*

Fifty partial paradigms (including infinitive, present participle, past participle, present indicative and präteritum forms) of German verbs were selected from the Celex database (Baayen, Piepenbrock, & Gulikers, 1995), totalling 752 uniformly-distributed verb forms. Of them, 694 forms were selected to be part of the training set. The remaining 58 forms, from 32 of the 50 selected paradigms, were kept for testing and classified for the morphological processes they undergo: circumfixation (as in the past participle forms *gefragt* and *gesehen*), stem alternation & suffixation (as with present/präteritum forms *beginnen/begannt*, *bleibe/blieben* and *denkst/dachten*), suffixation only (as in *mache/machst/machen*). All forms were encoded as strings of capitalised letters preceded by '#' and ended by '$', and administered to a *T MAP* one letter at a time, with re-entrant Hebbian connections being reset upon '#'. Umlauted characters were encoded as lower-case digraphs (e.g. '#HoeREN$' for *hören*) and the sharp s 'ß' as 'ss' (e.g. '#HEIssEN$' for h*eißen*). In both cases, pairs of lower-case letters are processed as one symbol. All letters were encoded as mutually orthogonal binary vectors.

Ten map instances with the same parameter setting (size = 35×35, *a* = 0.5, *b* = 1) were trained on the set of 694 forms over 100 epochs each, and tested independently on the remaining 58 forms for lexical recall. For each test word *K*, we sampled the corresponding integrated activation pattern on the Σ map. was copied onto the *T MAP* for the input word *K* to be read off. Per-word recall accuracy was then measured according to equation (8) above, with θ = 0.05.

Given an input word *K* of length n_K, we estimated the recalled character at time *t* as BMU(*t*) of *H*(*t*), with *H*(*t*) defined according to the following three equations, each representing a different recall strategy:

$$\text{(13)}\quad \begin{cases} H(t) = \alpha \cdot \widehat{H}_K(t) + \beta \cdot H_T(t) \\ H_T(t) = M \cdot H(t-1) \end{cases} \qquad \text{R1}$$

$$\text{(14)}\quad \begin{cases} H(t) = \alpha \cdot \widehat{H}_K(t) + \beta \cdot H_T(t) \\ H_T(t) = M \cdot H(t-1) + M \cdot H_T(t-1) \end{cases} \qquad \text{R2}$$

$$\text{(15)}\quad \begin{cases} H(t) = \alpha \cdot \widehat{H}_K(t) + \beta \cdot H_T(t) \\ H_T(t) = M \cdot H(t-1) + M \cdot \left(\sqrt{D} - \left\| 1_{N\times 1} \cdot W_{BMU(t-1)} - W \right\|\right) \end{cases} \qquad \text{R3}$$

for $t = 2, \ldots, n_K$.

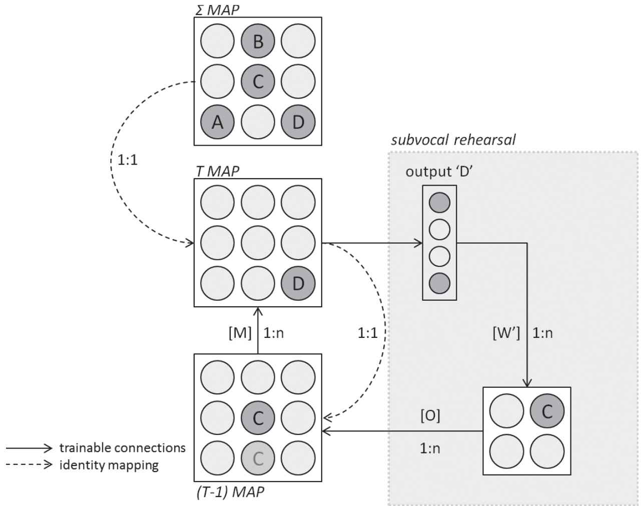


Figure 6. Lexical recall in a lexical tsom architecture augmented with a rehearsal component refreshing the auditory content of the *(t-1) map*. The *t-1* output 'C' is internally classified and input back to the *(t-1) map* for all 'C'-nodes to be reactivated. Expectations of 'C'-nodes are then used to output 'D' at time *t*.

Intuitively, *R1* (equation 13) defines *H(t)* as the integration of with the map's expectations at time *t*, based on *H(t-1)* according to equation (6) above and the lexical architecture of Figure 3. In *R2* (equation 14), the local expectation in (13) is augmented with the global expectation of the entire stored lexicon. This means that, in recalling a specific word *K*, a TSOM can take into account the range of expectations flowing from the word graph representing the entire memorised lexicon. Global expectations are time-aligned with the currently recalled character *bmu(t)*, and integrated with the local expectations prompted by the previously recalled character *bmu(t-1)*. Finally, *R3* (equation 15) is a variant of *R2*, whereby global expectations are selectively integrated with the expectations of all nodes recoding the symbol associated with *bmu(t-1)*. These expectations are prompted by the input symbol recalled at time *t-1* being covertly rehearsed and fed back to refresh the map's activation pattern. Figure 6 depicts an augmented lexical architecture where a re-entrant activation pattern is produced by a mechanism of subvocal rehearsal similar to the phonological loop classically invoked by Baddeley (1986) for refreshing the short-term buffer's auditory content.

4.2 *Results*

Average scores of recall routines *R1*, *R2* and *R3* on known forms (training set) were 97.04 (σ = 0.34), 98.29 (σ = 0.39), 98.66 (σ = 0.54) respectively, with a consistent but statistically non-significant advantage of the *R3* strategy over *R1* an *R2*. To assess the independent impact of different morphological processes on the generalisation bias of our architecture, recall scores were evaluated for each of the three process classes in the test set. Scores of recall accuracy on novel forms (test set), averaged over 10 map instances, are plotted in Figure 7 (top panel). The panel shows accuracy scores on circumfixation, stem alternation & suffixation, and suffixation only. Results are arranged vertically and highlighted with different shades of grey. Each shaded column groups three boxes, one for each recall strategy: *R1*, *R2* and *R3*. Difference in accuracy on circumfixation between *R3* and *R2* is statistically significant ($p < 0.001$). Likewise, difference in accuracy on stem alternation between *R2* and *R1* is equally significant ($p < 0.001$). Finally, the bottom panel of Figure 7 shows the average distance between *BMU* chains in both recoding and recall. The distance correlates negatively with recall accuracy scores. Recall is accurate when the map can restore, in recall, the activation chain that was produced in recoding.

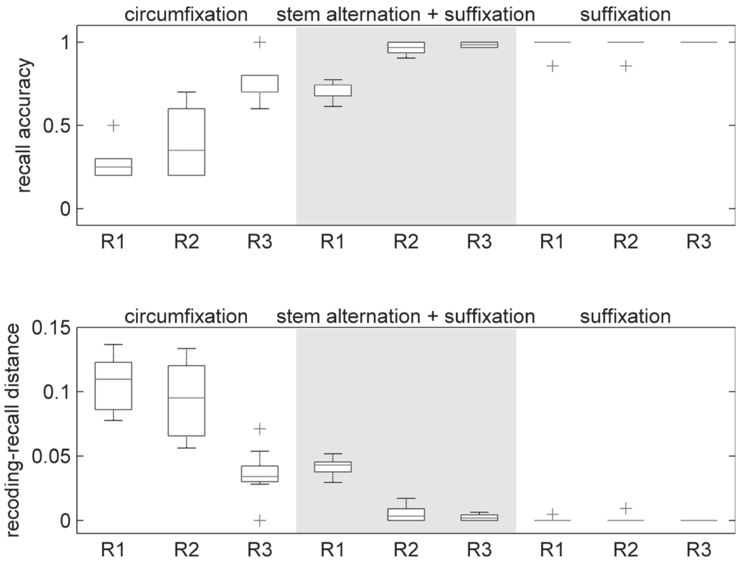


Figure 7. Top panel: boxplot distribution of accuracy scores on lexical recall of novel German forms (test set), plotted by morphological processes and recall strategies. Bottom panel: boxplot distribution of topological distance between recoding and recall activation chains for the same novel German forms.

5. GENERAL DISCUSSION

TSOMs define more LEVELS of input encoding. A peripherally encoded input vector is eventually RECODED on a map proper. It is the latter level that provides the long-term representation whereby the structure of input words is eventually PERCEIVED. In acquiring morphological structure, this is important for two reasons. First, it is neuro-biologically plausible, in keeping with what we know about other levels of representation of symbolic sequences in the human brain. For example, visual representations of written characters in the brain range from very concrete location-specific patterns of geometric lines in the occipito-temporal area of the left hemisphere, to more temporal representations which abstract away from physical features of letters such as position in the visual space, case and font (Dehaene, 2009). Secondly, TSOM's recoding is sensitive to input conditions. Since recoding is a function of training, cumulated activation patterns resulting from exposure to different data lead to different recoded representations. Other connectionist architectures, which fail to make a distinction between levels of input encoding, fall prey of *ad hoc* representational schemata such as Wickelcoding and positional coding, which are given at the outset and do not develop as a result of input exposure.

Evidence of German inflectional processes, ranging from prefixation and suffixation to stem alternation, shows that different alignment strategies may be required to acquire the same language. In our experiment, we assessed three such strategies on the task of recalling novel word forms.

R1 (equation 13) is based exclusively on topological propagation of temporal connections over neighbouring map nodes (Figure 2). The strategy enforces inter-node training dependence, allowing expectations of up-coming symbols to be transferred from one activation chain to another, based of topological contiguity (Figure 4, right). *R1* proves to be effective in generalising suffixal inflection, as shown in the top panel of Figure 7, but it fares disappointingly on past participles and stem alternations. In fact, a corollary of *R1* is that inflectional endings are better aligned if they are immediately preceded by the same left context. However, in generalising vowel alternation – say, from *binden* to *banden* based on the analogy to *finden* vs. *fanden* – the identity condition in the left context is not met.

R2 (equation 14) makes it up for the left-context bias by temporally aligning the word to be recalled with memorised alternating words of other paradigms. This makes the map considerably more proficient in generalising over stem alternating forms (Figure 7, top panel). However, time-aligned propagation of inter-paradigmatic forms performs poorly on past participles, the small improvement of *R2* over *R1* being statistically non-significant. Most German past participles require an intra-paradigmatic generalisation strategy, for a *ge-* prefixed stem (*ge-macht*) to be aligned with the corresponding stem in wordinitial position (*macht*). In other words, the map has to be tolerant to variation in time position of recurrent morphological constituents.

R3 (equation 15) addresses this issue through subvocal rehearsal (Figure 6). The input symbol -*m*- in *gemacht* re-enters the symbol-level map after recall, and activates other *m*-nodes, including those discharging in association with *macht* (and other non-prefixed forms of the same paradigm). Co-activation of paradigmatically-related forms through subvocal rehearsal puts the map in a stronger position to guess novel German past participles as well as stem alternating forms.

To sum up, recall of novel forms has to do with paradigm induction, and requires considerable flexibility in aligning the word to be recalled with other morphologically-related, stored word forms. Proper alignment is not based on local expectations only, but it requires that the analogical pressure of intra-paradigmatically and inter-paradigmatically related forms be brought to bear. At the same time, it relies on a proper, time-sensitive recoding of the symbols making up both stored and novel words. The extensive alignment between recoding activation chains and recall activation chains (Figure 7, bottom panel) show that alignment and recoding lie at the root of morphological generalisation, supporting the view that representation and processing issues are mutually implied in lexical competence. Finally, the success of *R3* highlights the importance of re-entrant feed-back mechanisms in monitoring generalisations.

6. CONCLUDING REMARKS

The computational analysis offered here accords well with recent neuro-physiological evidence of a bidirectional perisylvian pathway in the human brain, going from the superior temporal gyrus (Wernicke's area) to the Broca's area through the Inferior Parietal Lobule (Catani, Jones & ffytche, 2005). The pathway provides the neuro-cognitive substrate to the retention of sequences of linguistic units and orosensory goals for their vocalisation in working memory, lending support to the centrality of memory-based re-entrant mechanisms in language processing. This allows us to establish an interesting connection between issues of word representation, storage and processing on the one hand, and aspects of lexical architecture on the other hand.

Our computational analysis of low-level aspects of word recoding and recall is only a first step towards the development of a full-fledged lexical architecture based on memory self-organisation. It shows, nonetheless, that we can learn a lot about word structure by focusing on a range of issues that are normally taken for granted in the theoretical debate on morphology: from the ontogenesis of word representations to the dynamic of memory processes. By bringing together paradigm-based approaches to verb inflection and psycho-cognitive insights into the mental representation of time-series of symbols, we investigated the hypothesis that acquisition of verb inflection consists in the dynamic organisation of time-bound memory traces of fully inflected forms. Redundant morphological schemata emerge from self-organisation as time-aligned patterns of node activation shared by memorised forms. Notably, the schemata depend on both i) speakers' recoding and storage strategies, and ii) the underlying paradigm structure of the language to be acquired. This perspective on word acquisition and processing is in line with a construction-based reconceptualization of language rules as emergent lexical schemata (Jackendoff, 2002; Booij, 2010) and is conducive to a coherent ABSTRACTIVE model of morphological competence (Blevins, 2006; Pirrelli, Ferro & Marzi, forthcoming).

REFERENCES

Albright, A. (2002). Islands of reliability for regular morphology: evidence from Italian. *Language*, 78, 684-709.

Albright, A. & Hayes, B. (2002). Modeling English past tense intuitions with minimal generalization. In *Proceedings of the ACL 2002 Workshop on Morphological and Phonological Learning.* ACL Publications.

Baroni, M., Matiasek, J. & Trost, H. (2002). Unsupervised discovery of morphologically related words based on orthographic and semantic similarity. In *Proceedings of the Workshop on Morphological and Phonological Learning of ACL/SIGPHON-2002* (pp. 48-57). Philadelphia.

Baayen, H., Piepenbrock, R. & Gulikers, L. (1995). *The CELEX Lexical Database* (CD-ROM). Philadelphia: Linguistic Data Consortium.

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

Berger, A., Della Pietra, S. & Della Pietra, V. (1996). A maximum-entropy approach to Natural Language Processing. *Computational Linguistics* 22 (1), 39-71.

Bernhard, D. (2008). Simple morpheme labelling in unsupervised morpheme analysis. In C. Peters, V. Jijkoun, T. Mandl, H. Müller, D. W. Oard, A. Peñas, V. Petras, D. Santos (Eds.), *Advances in multilingual and multimodal information retrieval.* 8th Workshop of the Cross-Language Evaluation Forum, CLEF 2007 (Budapest, Hungary, September 19-21, 2007). Revised Selected Papers (pp. 873-880). Berlin: Springer.

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

Booij, G. (2010). *Construction Morphology*. Oxford: Oxford University Press.

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

Burzio, L. (2004). Paradigmatic and syntagmatic relations in italian verbal inflection. In Auger, J., Clements, J. C. & Vance, B. (Eds.), *Contemporary approaches to Romance linguistics.* Amsterdam/Philadelphia: Benjamins.

Bybee, J. (1995). Regular morphology and the lexicon. *Language and Cognitive Processes* 10 (5), 425-455.

Catani, M., Jones, D. K. & ffytche, D. H. (2005). Perisylvian language networks of the human brain. *Annals of Neurology* 57, 8–16.

Daelemans, W. & van den Bosch, A. (2005). Memory-based language processing. Cambridge: Cambridge University Press.

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.

Dehaene, S. (2009). *Reading the brain*. New York: Penguin.

Ferro, M., Marzi, C. & Pirrelli, V. (2011). A self-organizing model of word storage and processing: implications for morphology learning. *Lingue e Linguaggio* X (2), 209-226.

Freitag, D. (2005). Morphology induction from term clusters. In *Proceedings of the Ninth Conference on Computational Natural Language Learning (CoNLL-2005)* (pp. 128-135). Ann Arbor.

Goldsmith, J. (2001). Unsupervised learning of the morphology of natural language. *Computational Linguistics* 27 (2), 153-198.

Gaussier, É. (1999). Unsupervised learning of derivational morphology from inflectional lexicons. In *Proceedings of the Workshop on Unsupervised Learning in Natural Language Processing at the 37th Annual Meeting of the Association for Computational Linguistics (ACL-1999)* (pp. 24-30). Philadephia.

Golcher, F. (2006). Statistical text segmentation with partial structure analysis. In *Proceedings of KONVENS 2006* (pp. 44-51). Konstanz.

Goldsmith, J. (2001). Unsupervised learning of the morphology of a natural language. *Computational Linguistics* 27 (2), 153-198.

Hafer, M. A. & Weiss, S. F. (1974). Word segmentation by letter successor varieties. *Information Storage and Retrieval* 10, 371-385.

Harris, Zellig S. (1955). From phoneme to morpheme. *Language* 31 (2), 190-222.

Hammarström, H. (2009). *Unsupervised learning of morphology and the languages of the world.* Ph.D. dissertation, Chalmers University of Technology and University of Gothenburg.

Hammarström, H. & Borin, L. (2011). Unsupervised learning of morphology. *Computational Linguistics* 37 (2), 309-350.

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

Jackendoff, R. (2002), *Foundations of language. Brain, meaning, grammar, evolution*. Oxford: Oxford University Press.

Juola, P., Hall, C. & Boggs, A. (1994). Corpus-based morphological segmentation by entropy changes. In Monaghan, A. I. C. (Ed.), *Third Conference on the Cognitive Science of Natural Language Processing.* Dublin City University.

Lashley, K. (1951). The problem of serial order in behavior. In L. A. Jeffries (Ed.), *Cerebral mechanisms in behavior* (pp. 112-136). New York: John Wiley & Sons.

Marzi, C., Ferro, M., Caudai, C. & Pirrelli, V. (2012a). Evaluating Hebbian self-organizing memories for lexical representation and access. In *Proceedings of 8th International Conference on Language Resources and Evaluation (ELRA - LREC 2012)* (pp. 886-893).

Marzi, C., Ferro, M., & Pirrelli, V. (2012b). Prediction and generalisation in word processing and storage. In *8th Mediterranean Morphology Meeting Proceedings on Morphology and the architecture of ther grammar*, 113-130.

Matthews, P. H. (1991). *Morphology*. 2nd edition. Cambridge: Cambridge University Press.

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

Pirrelli, V. & Yvon, F. (1999). The hidden dimension: a paradigmatic view of data-driven NLP. *Journal of Experimental & Theroretical Artifical Intelligence* 11, 391-408.

Pirrelli, V. (2000). *Paradigmi in morfologia. Un approccio interdisciplinare alla flessione verbale dell'italiano*. Pisa-Roma: IEPI.

Pirrelli, V., Ferro, M. & Calderone, B. (2011). Learning paradigms in time and space. Computational evidence from Romance languages. In Maiden, M., Smith, J. C., Goldbach, M. & Hinzelin, M. O. (Eds.) *Morphological autonomy: perspectives from Romance inflectional morphology* (pp. 135-157). Oxford: Oxford University Press.

Pirrelli, V., Ferro, M. & Marzi, C. (forthcoming). Computational complexity of abstractive morphology. In Baerman, M., Brown, D. & Corbett, G. (Eds.), *Understanding and measuring morphological complexity*.

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.

Plunkett, K. & Juola, P. (1999). A connectionist model of English past tense and plural morphology. *Cognitive Science* 23 (4), 463-490.

Pollack, J. B. (1990). Recursive distributed representations. *Artificial Intelligence* 46, 77-105.

Ratnaparkhi, A. (1998). *Maximum entropy models for natural language ambiguity resolution*. Ph.D. dissertation, University of Pennsylvania.

Rodrigues, P. & Ćavar, D. (2005). Learning Arabic morphology using information theory. In *The Panels 2005: Proceedings from the Annual Meeting of the Chicago Linguistic Society.* Vol. 41 (2), 49-58, Chicago.

Rodrigues, P. & Ćavar, D. (2007). Learning Arabic morphology using statistical constraint-satisfaction models. In Benmamoun, E. (Ed.), *Perspectives on Arabic linguistics: papers from the Annual Symposium on Arabic Linguistics* (pp. 63-75). Urbana.

Rumelhart, D. & McClelland, J. (1986), On learning the past tense of English verbs. In D. E. Rumelhart & J. McClelland (Eds.), *Parallel distributed processing: explanations in the microstructure of cognition*. Cambridge: The MIT Press.

Schone, P. & Jurafsky, D. (2000). Knowledge-free induction of inflectional morphologies using latent semantic analysis. In *Conference on Natural Language Learning 2000 (CoNLL-2000)* (pp. 67-72). Lisbon.

Schone, P. & Jurafsky, D. (2001). Is knowledge-free induction of multiword unit dictionary headwords a solved problem? In *Proceedings of Empirical Methods in Natural Language Processing*, Pittsburgh.

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.

Stump, G. T. (2001). *Inflectional morphology. a theory of paradigm structure*. Cambridge: Cambridge University Press.

Uchimoto, K., Sekine, S. & Isahara, H. (2001). The unknown word problem: a morphological analysis of Japanese using maximum entropy aided by a dictionary. In *Proceedings of Empirical Methods in Natural Language Processing*, 91-99.

Wickelgren, W. A. (1965). Acoustic similarity and intrusion errors in short-term memory. *Journal of Experimental Psychology* 70, 102-108.

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

Xanthos, A. (2007). *Apprentissage automatique de la morphologie: le cas des structures racine-schème*. Ph.D. dissertation, Université de Lausanne.

Yarowsky, D. & Wicentowski, R. (2000). Minimally supervised morphological analysis by multimodal alignment. In *Proceedings of the 38th Annual Meeting of the Association for Computational Linguistics (ACL-2000)* (pp. 207-216). Hong Kong.

Claudia Marzi
Istituto di Linguistica Computazionale “A. Zampolli”
Consiglio Nazionale delle Ricerche
Via G. Moruzzi 1 – 56124 Pisa
Italy
e-mail: claudia.marzi@ilc.cnr.it

Marcello Ferro
Istituto di Linguistica Computazionale “A. Zampolli”
Consiglio Nazionale delle Ricerche
Via G. Moruzzi 1 – 56124 Pisa
Italy
e-mail: marcello.ferro@ilc.cnr.it

Vito Pirrelli
Istituto di Linguistica Computazionale “A. Zampolli”
Consiglio Nazionale delle Ricerche
Via G. Moruzzi 1 – 56124 Pisa
Italy
e-mail: vito.pirrelli@ilc.cnr.it