Proceedings of the Workshop of the
ACL Special Interest Group on Computational Phonology (SIGPHON)
Association for Computations Linguistics
Barcelona, July 2004

Non-locality all the way through:
Emergent Global Constraints in the Italian Morphological Lexicon

Vito Pirrelli
Istituto di Linguistica Computazionale
CNR, Pisa, Italy
vito.pirrelli@ilc.cnr.it

Basilio Calderone
Laboratorio di Linguistica
Scuola Normale Superiore, Pisa, Italy
b.calderone@sns.it

Ivan Herreros
Istituto di Linguistica Computazionale
CNR, Pisa, Italy
ivan.herreros@ilc.cnr.it

Michele Virgilio
Dipartimento di Fisica
Università degli Studi di Pisa, Italy
virgilio@df.unipi.it

Abstract

The paper reports on the behaviour of a Kohonen map of the mental lexicon, monitored through different phases of acquisition of the Italian verb system. Reported experiments appear to consistently reproduce emergent *global* ordering constraints on memory traces of inflected verb forms, developed through principles of *local interactions* between parallel processing neurons.

1 Introduction

Over the last 15 years, considerable evidence has accrued on the critical role of paradigm-based relations as an order-principle imposing a non-local organising structure on word forms memorised in the speaker's mental lexicon, facilitating their retention, accessibility and use, while permitting the spontaneous production and analysis of novel words. A number of theoretical models of the mental lexicon have been put forward to deal with the role of these global constraints in i) setting an upper bound on the number of possible forms a speaker is ready to produce (Stemberger and Carstairs, 1988), ii) accounting for reaction times in lexical decision and related tasks (Baayen *et al.* 1997; Orsolini and Marslen-Wilson, 1997 and others), iii) explaining production errors by both adults and children (Bybee and Slobin, 1982; Bybee and Moder; 1983; Orsolini *et al.*, 1998) and iv) accounting for human acceptability judgements and generalisations over nonce verb stems (Say and Clahsen, 2001). While most of these models share some core assumptions, they appear to largely differ on the role played by lexical relations in word storage, access and processing. According to the classical view (e.g. Taft, 1988) the relationship between regularly inflected forms is directly encoded as lexical procedures linking inflectional affixation to separately encoded lexical roots. Irregular word forms, on the other hand, are stored in full (Prasada and Pinker, 1993). In contrast to this view, associative models of morphological processing claim that words in the mental lexicon are always listed as full forms, establishing an interconnected network of largely redundant linguistic data reflecting similarities in meaning and form (Bybeee, 1995).

Despite the great deal of experimental evidence now available, however, we still seem to know too little of the dynamic interplay between morphological learning and the actual working of the speaker's lexicon to draw conclusive inferences from experimental findings. Associative models, for example, are generally purported to be unable to capture morpheme-based effects of morphological storage and access. Thus, if humans are shown to access the mental lexicon through morphemes, so the argument goes, then associative models of the mental lexicon cannot be true. In fact, if associative models *can* simulate emergent morpheme-based effects of lexical organisation through storage of full forms, then this conclusion is simply unwarranted.

We believe that computer simulations of morphology learning can play a role in this dispute. However, there have been comparatively few attempts to model the way global ordering principles of lexical organisation interact with (local) processing strategies in morphology learning. In the present paper, we intend to simulate a biologically-inspired process of paradigm-based self-organisation of inflected verb forms in a Kohonen map of the Italian mental lexicon, built on the basis of local processes of memory access and updating. Before we go into that, we briefly overview relevant machine learning work from this perspective.

2 Background

Lazy learning methods such as the *nearest neighbour algorithm* (van den Bosch et al., 1996) or the *analogy-based* approach (Pirrelli and Federici, 1994; Pirrelli and Yvon, 1999) require full storage of supervised data, and make on-line use of them with no prior or posterior lexical structuring. This makes this class of algorithms flexible and efficient, but comparatively noise-sensitive and rather poor in simulating emergent learning phenomena. There is no explicit sense in which the system learns how to map new exemplars to already memorised ones, since the mapping function does not change through time and the only incremental pay-off lies in the growing quantity of information stored in the exemplar data-base.

Decision tree algorithms (Quinlan, 1986), on the other hand, try to build the shortest hierarchical structure that best classifies the training data, using a greedy heuristics to select the most discriminative attributes near the root of the hierarchy. As heuristics are based on a locally optimal splitting of all training data, adding new training data may lead to a dramatic reorganisation of the hierarchy, and nothing is explicitly learned from having built a decision tree at a previous learning stage (Ling and Marinov, 1993).

To tackle the issue of word structure more squarely, there has been a recent upsurge of interest in global *paradigm-based constraints* on morphology learning, as a way to minimise the range of inflectional or derivational endings heuristically inferred from raw training data (Goldsmith, 2001; Gaussier, 1999; Baroni, 2000). It should be noted, however, that global, linguistically-inspired constraints of this sort do not interact with morphology learning in any direct way. Rather, they are typically used as global criteria for optimal convergence on an existing repertoire of minimally redundant sets of paradigmatically related morphemes. Candidate morpheme-like units are acquired independently of paradigm-based constraints, solely on the basis of local heuristics. Once more, there is no clear sense in which global constraints form integral part of learning.

Of late, considerable attention has been paid to aspects of emergent morphological structure and continuous compositionality in *multi-layered perceptrons*. Plaut et al. (1996) show how a neural network comes to be sensitive to degrees of compositionality on the basis of exposure to examples of inputs and outputs from a word-reading task. Systematic input-output pairs tend to establish a clear one-to-one correlation between parts of input and parts of output representations, thus developing strongly compositional analyses. By the same token, a network trained on inputs with graded morphological structure develops representations with corresponding degrees of compositionality (Rueckl and Raveh, 1999). It must be appreciated that most such approaches to incremental compostionality are task-oriented and highly supervised. Arguably, a better-motivated and more explanatory approach should be based on self-organisation of input tokens into morphologically natural classes and their time-bound specialisation as members of one such class, with no external supervision. Kohonen's Self-Organising Maps (*SOM*s) (Kohonen, 1995) simulate self-organisation by structuring input knowledge on a (generally) two-dimensional grid of neurons, whose activation values can be inspected by the researcher both instantaneously and through time. In the remainder of this paper we show that we can use SOMs to highlight interesting aspects of global morphological organisation in the learning of Italian conjugation, incrementally developed through local interactions between parallel processing neurons.

3 SOMs

SOMs can project input tokens, represented as data points of an *n*-dimensional *input space*, onto a generally two-dimensional *output space* (the map grid) where similar input tokens are mapped onto nearby output units. Each output unit in the map is associated with a distinct *prototype vector*, whose dimensionality is equal to the dimensionality of input vectors. As we shall see, a prototype vector is an approximate memory trace of recurring inputs, and plays the role of linking its corresponding output unit to a position in the input space. Accordingly, each output unit takes two positions: one in the input space (through its prototype vector) and one in the output space (its co-ordinates on the map grid).

SOMs were originally conceived of as computer models of somatotopic *brain maps*. This explains why output units are also traditionally referred to as *neurons*. Intuitively, a prototype vector represents the memorised input pattern to which its associated neuron is most sensitive. Through learning, neurons gradually specialise in selectively being associated with specific input patterns. Moreover, memorised input patterns tend to cluster on the map grid so as to reflect natural classes in the input space.

These interesting results are obtained through iterative unsupervised exposure to input tokens. At each learning step, a SOM is exposed to a single input token and goes through the following two stages: a) competitive neuron selection, and b) adaptive adjustment of prototype vectors. As we shall see in more detail in the remainder of this section, both stages are local and incremental in some crucial respects.[1]

3.1 Stage 1: competitive selection

Let v_x be the *n*-dimension vector representation of the current input. At this stage, the distance between each prototype vector and v_x is computed. The output unit *b* that happens to be associated with the prototype vector v_b closest to v_x is selected as the *best matching unit*. More formally:

$$\|v_x - v_b\| \equiv \min\{\|v_x - v_i\|\},$$

where is also known as the *quantization error* scored by v_b relative to v_x. Intuitively, this is to say that, although *b* is the map neuron reacting most sensitively to the current stimulus, *b* is not (yet) perfectly attuned to v_x.

Notably, the quantization error is a local distance function, as it involves two vector representations at a time. Hence, competitive selection is blind to general structural properties of the input space, such as the comparative role of each dimension in discriminating input tokens. This makes competitive selection prone to errors due to accidental or spurious similarity between the input vector and SOM prototype vectors.

3.2 Stage 2: adaptive adjustment

After the winner unit *b* is selected at time *t*, the SOM locally adapts prototype vectors to the current stimulus. Vector adaptation applies locally, within a kernel area of radius *r*, centred on the position of *b* on the map grid. Both $v_b(t)$ (v_b at time *t*) and the prototype vectors associated with *b*'s kernel units are adjusted to make them more similar to $v_x(t)$ (v_x at time *t*). In particular, for each prototype vector v_i in *b*'s kernel and the input vector v_x, the following adaptive function is used

$$v_i(t+1) = v_i(t) + h_{bi}\left[v_x(t) - v_i(t)\right] ,$$

where h_{bi} is the neighbourhood kernel centred around the winner unit *b* at time *t*, a non-increasing function of both time and the distance between the input vi and the winner vector vb. As learning time progresses, however, h_{bi} decreases, and prototype vector updates become less sensitive to input conditions, according to the following:

$$h_{bi}(t) = h(\|l_b - l_i\|, t) \cdot \alpha(t) ,$$

where l_b and l_i are, respectively, the position of *b* and its kernel neurons on the map grid, and $\alpha(t)$ is the learning rate at time *t*, a monotonically decreasing function of *t*. Interaction of these functions simulates effects of memory entrenchment and proto-typicality of early input data.

3.3 Summary

The dynamic interplay between locality and incrementality makes SOMs plausible models of neural computation and data compression. Their sensitivity to frequency effects in the distribution of input data allows the researcher to carefully test their learning behaviour in different time-bound conditions. Learning makes output units increasingly more reactive to already experienced stimuli and thus gradually more competitive for selection. If an output unit is repeatedly selected by systematically occurring input tokens, it becomes associated with a more and more faithful vector representation of a stimulus or class of stimuli, to become an *attractor* for its neighbouring area on the map. As a result, the most parsimonious global organisation of input data emerges that is compatible with a) the size of the map grid, b) the dimensionality of output units and c) the distribution of input data.

This intriguing dynamics persuaded us to use SOMs to simulate the emergence of non-local lexical constraints from local patterns of interconnectivity between vector representations of full word forms. The Italian verb system offers a particularly rich material to put this hypothesis to the challenging test of a computer simulation.

4 The Italian Verb System

The Italian conjugation is a complex inflectional system, with a considerable number of classes of regular, subregular and irregular verbs exhibiting different probability densities (Pirrelli, 2000; Pirrelli and Battista, 2000). Traditional descriptive grammars (e.g. Serianni, 1988) identify three main conjugation classes (or more simply conjugations), characterised by a distinct thematic vowel (TV), which appears between the verb root and the inflectional endings. First conjugation verbs have the TV *-a-* (*parl-a-re* 'speak'), second conjugation verbs have the TV *-e-* (*tem-e-re* 'fear'), and third conjugation verbs *-i-* (*dorm-i-re* 'sleep'). The first conjugation is by far the largest class of verbs

[1] This marks a notable difference between SOMs and other classical projection techniques such as Vector Analysis or Multi-dimensional Scaling, which typically work on the basis of global constraints on the overall distribution of input data (*e.g.* by finding the space projection that maximizes data variance/co-variance).

TYPE	EXAMPLE	ENGLISH GLOSS
[isk]-insertion + palatalization	fi'nisko/fi'niʃʃi/fi'njamo	(I)/(you)/(we) end
[g]-insertion + diphthongization	'vɛngo/'vjɛni/ve'njamo	(I)/(you)/(we) come
ablauting + velar palatalization	'ɛsko/'ɛʃʃi/uʃ'ʃamo	(I)/(you)/(we) go out
[r]-drop + diphthongization	'mwojo/'mwori/mo'rjamo	(I)/(you)/(we) die

Table 1. Variable stem alternations in the Italian present indicative.

(73% of all verbs listed in De Mauro et al., 1993), almost all of which are regular. Only very few 1st conjugation verbs have irregularly inflected verb forms: *andare* 'go', *dare* 'give', *stare* 'stay' and *fare* 'do, make'. It is also the only truly productive class. Neologisms and foreign loan words all fall into it. The second conjugation has far fewer members (17%), which are for the most part irregular (around 95%). The third conjugation is the smallest class (10%). It is mostly regular (around 10% of its verbs are irregular) and only partially productive.

Besides this macro-level of paradigmatic organisation, Italian subregular verbs also exhibit ubiquitous patterns of stem alternations, whereby a change in paradigm slot triggers a simultaneous change of verb stem and inflectional ending, as illustrated in Table 1 for the present indicative active. Pirrelli and Battista (2000) show that phenomena of Italian stem alternation, far from being accidental inconsistencies of the Italian morphophonology, define stable and strikingly convergent patterns of variable stem formation (Aronoff, 1994) throughout the entire verb system. The patterns partition subregular Italian verbs into equivalence micro-classes. In turn, this can be interpreted as suggesting that inter-class consistency plays a role in learning and may have exerted a convergent pressure in the history of the Italian verb system. If a speaker has heard a verb only in ambiguous inflections (i.e. inflections that are indicators of more than one verb micro-class), (s)he will need to guess, in order to produce unambiguous forms. Guesses are made on the basis of frequently attested verb micro-classes (Albright, 2002).

5 Computer simulations

The present experiments were carried out using the SOM toolbox (Vesanto *et al.*, 2000), developed at the Neural Networks Research Centre of Helsinki University of Technology. The toolbox partly forced some standard choices in the training protocol, as discussed in more detail in the following sections. In particular, we complied with Kohonen's view of SOM training as consisting of two successive phases: a) *rough training* and b) *fine-tuning*. The implications of this view will be discussed in more detail later in the paper.

5.1 Input data

Our input data are inflected verb forms written in standard Italian orthography. Since Italian orthography is, with a handful of exceptions, consistently phonological, we expect to replicate the same results with phonologically transcribed verb forms.

Forms are incrementally sampled from a training data set, according to their probability densities in a free text corpus of about 3 million words. Input data cover a fragment of Italian verb inflection, including, among others, present indicative active, future indicative active, infinitive and past participle forms, for a total of 10 different inflections. The average length of training forms is 8.5, with a max value of 18.

Following Plunkett and Marchman (1993), we assume than the map is exposed to a gradually growing lexicon. At epoch 1, the map learns inflected forms of the 5 most frequent verb types. At each ensuing epoch, five more verb types are added to the training data, according to their rank in a list of decreasingly frequent verb types. As an overall learning session consists of 100 epochs, the map is eventually exposed to a lexicon of 500 verb types, each seen in ten different inflections. Although forms are sampled according to their corpus distributions, we hypothesise that the range of inflections in which verb tokens are seen by the map remains identical across verb types. This is done to throw paradigmatic effects in sharper relief and responds to the (admittedly simplistic) assumption that the syntactic patterns forming the linguistic input to the child do not vary across verb types.

Each input token is localistically encoded as an 8*16 matrix of values drawn from the set {1, -1}. Column vectors represent characters, and rows give the random encoding of each character, ensuring maximum independence of character vector representations. The first eight columns in the matrix represent the first left-aligned characters of the form in question. The remaining eight columns stand for the eight (right-aligned) final characters of the input form.

a) b)

Figure 1. Early self-organisation of a SOM for roots (a) and endings (b) of Italian verbs (epoch 10).

a) b)

Figure 2. Late self-organization of a SOM for roots (a) and endings (b) of Italian verbs (epoch 100).

5.2 Training protocol

At each training epoch, the map is exposed to a total of 3000 input tokens. As the range of different inflected forms from which input tokens are sampled is fairly limited (especially at early epochs), forms are repeatedly shown to the map. Following Kohonen (1995), a learning epoch consists of two phases. In the first rough training phase, the SOM is exposed to the first 1500 tokens. In this phase, values of α (the learning rate) and neighbourhood kernel radius r are made vary as a linear decreasing function of the time epoch, from max $\alpha = 0.1$ and $r = 20$ (epoch 1), to $\alpha = 0.02$ and $r = 10$ (epoch 100). In the second fine-tuning phase of each epoch, on the other hand, α is kept to 0.02 and $r = 3$.

5.3 Simulation 1: Critical transitions in lexical organisation

Figures 1 and 2 contain snapshots of the Italian verb map taken at the beginning and the end of training (epochs 1 and 100). The snapshots are Unified distance matrix (U-matrix, Ultsch and Siemon, 1990) representations of the Italian SOM. They are used to visualise distances between neurons. In a U-matrix representation, the distance between adjacent neurons is calculated and presented with different colourings between adjacent positions on the map. A dark colouring between neurons signifies that their corresponding prototype vectors are close to each other in the input space. Dark colourings thus highlight areas of the map whose units react consistently to the same stimuli. A light colouring between output units, on the other hand, corresponds to a large distance (a gap) between their corresponding prototype vectors. In short, dark areas can be viewed as clusters, and light areas as chaotically reacting cluster separators. This type of pictorial presentation is useful when one wants to inspect the state of knowledge developed by the map through learning.

For each epoch, we took two such snapshots: i) one of prototype vector dimensions representing the *initial part* of a verb form (approximately its verb root, Figures 1.a and 2.a), and ii) one of prototype vector dimensions representing the verb *final part* (approximately, its inflectional endings, Figure 1.b and 2.b).

5.3.1 Discussion

Data storage on a Kohonen map is a dynamic process whereby i) output units tend to consistently become more reactive to classes of input data, and ii) vector prototypes which are adjacent in the input space tend to cluster in topologically connected subareas of the map.

Self-organisation is thus an emergent property, based on local (both in time and space) principles of prototype vector adaptation. At the outset, the map is a *tabula rasa*, *i.e.* it has no notion whatsoever of Italian inflectional morphology. This has two implications. First, before training sets in, output units are associated with randomly initialised sequences of characters. Secondly, prototype vectors are randomly associated with map neurons, so that two contiguous neurons on the map may be sensitive to very different stimulus patterns.

Figure 1 shows that, after the first training epoch, the map started by organising memorised input patterns lexically, grouping them around their (5) roots. Each root is an attractor of lexically related stimuli, that nonetheless exhibit fairly heterogeneous endings (see Figure 1.b).

At learning epoch 100, on the other hand, the topological organisation of the verb map is the mirror image of that at epoch 10 (Figures 2.a and 2.b). In the course of learning, root attractors are gradually replaced by ending attractors. Accordingly, vector prototypes that used to cluster around their lexical root appear now to stick together by morpho-syntactic categories such as tense, person and number. One can conceive of each connected dark area of map 2.b as a slot in an abstract inflectional paradigm, potentially associated with many forms that share an inflectional ending but differ in their roots.

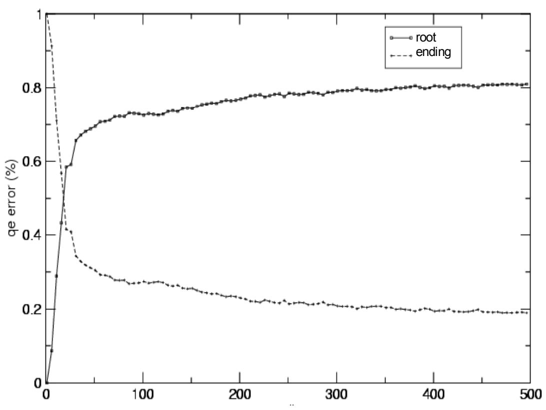

Figure 3. Average quantization error for an increasing number of input verbs

The main reason for this morphological organisation to emerge at a late learning stage rests in the distribution of training data. At the beginning, the map is exposed to a small set of verbs, each of which is inflected in 10 different forms. Forms with the same ending tend to be fewer than forms with the same root. As the verb vocabulary grows (say of the order of about 50 different verbs), however, the principles of morphological (as opposed to lexical) organisation allow for more compact and faithful data storage, as reflected by a significant reduction in the map average quantization error (Figure 3). Many different forms can be clustered around comparatively few endings, and the latter eventually win out as local paradigmatic attractors.

Figure 4 (overleaf) is a blow-up of the map area associated with infinitive and past participle endings. The map shows the content of the last three characters of each prototype vector. Since past participle forms occur in free texts more often than infinitives, they have a tendency to take a proportionally larger area of the map (due to the so-called *magnification factor*). Interestingly enough, past participles ending in *-ato* occupy one third of the whole picture, witnessing the prominent role played by regular first conjugation verbs in the past participle inflection.

Another intriguing feature of the map is the way the comparatively connected area of the past participle is carved out into tightly interconnected micro-areas, corresponding to subregular verb forms (e.g. *corso* 'run', *scosso* 'shaken' and *chiesto* 'answered'). Rather than lying outside of the morpho-phonological realm (as exceptions to the "TV + *to*" default rule), subregular forms of this kind seem here to draw the topological borders of the past participle domain, thus defining a continuous chain of morphological family resemblances. Finally, by analogy-based continuity, the map comes to develop a prototype vector for the non existing (but paradigmatically consistent) past participle ending *-eto*.[2] This "spontaneous" overgeneralization is the by-product of graded, overlapping morpheme-based memory traces.

In general, stem frequency may have had a retardatory effect on the critical transition from a lexical to a paradigm-based organisation. For the same reason, high-frequency forms are eventually memorised as whole words, as they can successfully counteract the root blurring effect produced by the chaotic overlay of past participle forms of different verbs, which are eventually attracted to the same map area. This turns out to be the case for very frequent past participles such as *stato* 'been' and *fatto* 'done'. As a final point, a more detailed analysis of memory traces in the past participle area of the map is likely to highlight significant stem patterns in the subregular microclasses. If confirmed, this should provide fresh evidence supporting the existence of prototypical morphonological stem patterns consistently selecting specific subregular endings (Albright, 2002).

5.4 Simulation 2: Second level map

A SOM projects *n*-dimensional data points onto grid units of reduced dimensionality (usually 2). We can take advantage of this data compression to train a new SOM with complex representations consisting of the output units of a previously trained SOM. The newly trained SOM is a second level projection of the original data points.

To test the consistency of the paradigm-based organisation of the map in Figure 2, we trained a

[2] While Italian regular 1st and 3rd conjugation verbs present a thematic vowel in their past participle endings (*-ato* and *-ito* respectively), regular 2 conjugation past participles (TV *-e-*) end, somewhat unexpectedly, in *-uto*.

novel SOM with verb type vectors. Each such vector contains all 10 inflected forms of the same verb type, encoded through the co-ordinates of their best-matching units in the map grid of Figure 2. The result of the newly trained map is given in Figure 5.

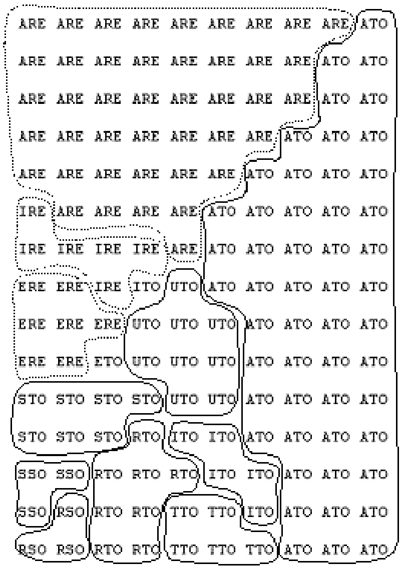

Figure 4. The past participle and infinitive areas

5.4.1 Discussion

Figure 5 consistently pictures the three-fold macrostructure of the Italian verb system (section 2) as three main horizontal areas going across the map top-to-bottom.

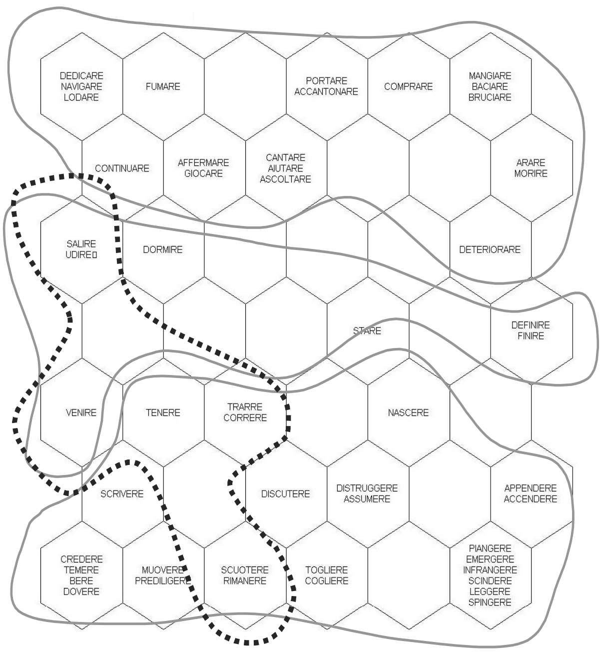

Figure 5: A second level map

Besides, we can identify other micro-areas, somewhat orthogonal to the main ones.The most significant such micro-class (circled by a dotted line) contains so-called [g]-inserted verbs (Pirrelli, 2000; Fanciullo, 1998), whose forms exhibit a characteristic [g]/0 stem alternation, as in vengo/venite 'I come, you come (plur.)' and tengo/tenete 'I have/keep, you have/keep (plur.)'. The class straddles the 2nd and 3rd conjugation areas, thus pointing to a convergent phenomenon affecting a portion of the verb system (the present indicative and subjunctive) where the distinction between 2nd and 3rd conjugation inflections is considerably (but not completely) blurred. All in all, Italian verbs appear to fall not only into equivalence classes based on the selection of inflectional endings (traditional conjugations), but also into homogeneous micro-classes reflecting processes of variable stem formation. Identification of the appropriate micro-class is a crucial problem in Italian morphology learning. Our map appears to be in a position to tackle it reliably.

Note finally the very particular position of the verb *stare* 'stay' on the grid. Although *stare* is a 1st conjugation verb, it selects some 2nd conjugation endings (e.g. *stessimo* 'that we stayed (subj.)' and *stette* '(s)he stayed'). This is captured in the map, where the verb is located halfway between the 1st and 2nd conjugation areas.

6 Conclusion and future work

The paper offered a series of snapshots of the dynamic behaviour of a Kohonen map of the mental lexicon taken in different phases of acquisition of the Italian verb system. The snapshots consistently portray the emergence of global ordering constraints on memory traces of inflected verb forms, at different levels of linguistic granularity.

Our simulations highlight not only morphologically natural classes of input patterns (reminiscent of the hierarchical clustering of perceptron input units on the basis of their hidden layer activation values) and selective specialisation of neurons and prototype vector dimensions in the map, but also other non-trivial aspects of memory organisation. We observe that the number of neighbouring units involved in the memorisation of a specific morphological class is proportional to both type frequency of the class and token frequency of its members. Token frequency also affects the entrenchment of memory areas devoted to storing individual forms, so that highly frequent forms are memorised in full, rather than forming part of a morphological cluster.

In our view, the solid neuro-physiological basis of SOMs' processing strategies and the considerable psycho-linguistic and linguistic evidence in favour of global constraints in morphology learning make the suggested approach an interesting medium-scale experimental framework, mediating between small-scale neurological structures and large-scale linguistic evidence. In the end, it would not be surprising if more in-depth computational analyses of this sort will give strong indications that associative models of the morphological lexicon are compatible with a "realistic" interpretation of morpheme-based decomposition and access of inflected forms in the mental lexicon. According to this view, morphemes appear to play a truly active role in lexical indexing, as they acquire an increasingly dominant position as local attractors through learning. This may sound trivial to the psycholinguistic community. Nonetheless, only very few computer simulations of morphology learning have so far laid emphasis on the importance of incrementally acquiring structure from morphological data (as opposed – say – to simply memorising more and more input examples) and on the role of acquired structure in lexical organisation. Most notably for our present concerns, the global ordering constraints imposed by morphological structure in a SOM are the by-product of purely local strategies of memory access, processing and updating, which are entirely compatible with associative models of morphological learning. After all, the learning child is not a linguist and it has no privileged perspective on all relevant data. It would nonetheless be somewhat reassuring to observe that its generalisations and ordering constraints come very close to a linguist's ontology.

The present work also shows some possible limitations of classical SOM architectures. The propensity of SOMs to fully memorise input data only at late learning stages (in the fine-tuning phase) is not fully justified in our context. Likewise, the hypothesis of a two-staged learning process, marked by a sharp discontinuity at the level of kernel radius length, has little psycholinguistic support. Furthermore, multiple classifications are only minimally supported by SOMs. As we saw, a paradigm-based organisation actually replaces the original lexical structure. This is not entirely desirable when we deal with complex language tasks. In order to tackle these potential problems, the following changes are currently being implemented:

- endogenous modification of radius length as a function of the local distance between the best matching prototype vector and the current stimulus; the smaller the distance the smaller the effect of adaptive updating on neighbouring vectors
- adaptive vector-distance function; as a neuron becomes more sensitive to an input pattern, it also develops a sensitivity to specific input dimensions; differential sensitivity, however, is presently not taken into account when measuring the distance between two vectors; we suggest weighting vector dimensions, so that distances on some dimensions are valued higher than distances on other dimensions
- "self-feeding" SOMs for multiple classification tasks; when an incoming stimulus has been matched by the winner unit only partially, the non matching part of the same stimulus is fed back to the map; this is intended to allow "recognition" of more than one morpheme within the same input form
- more natural input representations, addressing the issue of time and space-invariant features in character sequences.

References

Albright, Adam. 2002. Islands of reliability for regular morphology: Evidence from Italian. *Language*, 78:684-709.

Aronoff, Mark. 1994. *Morphology by Itself.* M.I.T. Press, Cambridge, USA.

Baayen, Harald, Ton Dijkstra and Robert Schreuder. 1997. Singulars and Plurals in Dutch: Evidence for a Parallel Dual Route Model. *Journal of Memory and Language*, 36:94-117.

Baroni, Marco. 2000. *Distributional cues in morpheme discovery: A computational model and empirical evidence*. Ph.D. dissertation, UCLA.

Bosch van den, Antal, Walter Daelemans, Ton Weijters. 1996. Morphological Analysis as Classification: an Inductive-learning approach. In *Proceedings of NEMLAP II* , K. Oflazer and H. Somers, eds., pages 79-89, Ankara.

Bybee, Joan. 1995. Regular Morphology and the Lexicon. *Language and Cognitive Processes*, 10 (5): 425-455.

Bybee, Joan and Dan I. Slobin. 1982. Rules and Schemas in the Development and Use of the English Past Tense. *Language*, 58:265-289.

Bybee, Joan and Carol Lynn Moder. 1983. Morphological Classes as Natural Categories. *Language*, 59:251-270.

De Mauro, Tullio, Federico Mancini, Massimo Vedovelli and Miriam Voghera. 1993. *Lessico di frequenza dell'italiano parlato*. Etas Libri, Milan.

Fanciullo, Franco. 1998. Per una interpretazione dei verbi italiani a "inserto" velare. *Archivio Glottologico Italiano*, LXXXIII(II):188-239.

Gaussier, Eric. 1999. Unsupervised learning of derivational morphology from inflectional lexicons. In *Proceedings of the Workshop on Unsupervised Learning in Natural Language Processing*, pages 24-30, University of Maryland.

Goldsmith, John. 2001. Unsupervised Learning of the Morphology of a Natural Language. *Computational Linguistics*, 27(2):153-198.

Kohonen, Teuvo. 1995. *Self-Organizing Maps*. Springer, Berlin.

Ling, Charles X. and Marin Marinov. 1993. Answering the Connectionist Challenge: a Symbolic Model of Learning the Past Tense of English Verbs. *Cognition*, 49(3):235-290.

Orsolini, Margherita and William Marslen-Wilson. 1997. Universals in Morphological Representations: Evidence from Italian. *Language and Cognitive Processes*, 12(1):1-47.

Orsolini, Margherita, Rachele Fanari and Hugo Bowles. 1998. Acquiring regular and irregular inflection in a language with verbal classes. *Language and Cognitive Processes*, 13(4):452-464.

Pirrelli, Vito. 2000. *Paradigmi in Morfologia.* Istituti Editoriali e Poligrafici Internazionali, Pisa.

Pirrelli, Vito and Federici Stefano. 1994. "Derivational" Paradigms in Morphonology. In *Proceedings of Coling 94*, pages 234-240, Kyoto.

Pirrelli, Vito and François Yvon. 1999. The hidden dimension: a paradigmatic view of data driven NLP. *Journal of Experimental and Theoretical Artificial Intelligence*, 11:391-408.

Pirrelli, Vito and Marco Battista. 2000. The Paradigmatic Dimension of Stem Allomorphy in Italian Inflection. *Italian Journal of Linguistics*, 12(2):307-380.

Plaut, David C., James L. McClelland, Mark S. Seidenberg and Karalyn Patterson. 1996. Understanding Normal and Impaired Word Reading: Computational Principles in Quasi-regular Domains. *Psychological Review* , 103:56-115.

Plunkett, Kim and Virginia Marchman. 1993. From rote learning to system building: Acquiring verb morphology in children and connectionist nets. *Cognition*, 48:21-69.

Prasada, Sandeep and Steven Pinker. 1993. Generalizations of regular and irregular morphology. Language and Cognitive Processes, 8:1-56.

Rueckl, Jay G. and Michal Raveh. 1999. The Influence of Morphological Regularities on the Dynamics of Connectionist Networks. *Brain and Language*, 68:110-117.

Say, Tessa and Harald Clahsen. 2001. *Words, Rules and Stems in the Italian Mental Lexicon*. In “Storage and computation in the language faculty”, S. Nooteboom, F. Weerman and F. Wijnen, eds., pages 75-108, Kluwer Academic Publishers, Dordrecht.

Serianni, Luca. 1988. *Grammatica italiana: italiano comune e lingua letteraria*. UTET, Turin.

Stemberger, Joseph P. and Andrew Carstairs. 1988. A Processing Constraint on Inflectional Homonymy. *Linguistics*, 26:601-61.

Taft, Marcus. 1988. A morphological-decomposition model of lexical representation. *Linguistics, 26*:657-667.

Ultsch, Alfred and H. Peter Siemon. 1990. Kohonen's Self-Organizing Feature Maps for Exploratory Data Analysis. In “Proceedings of INNC'90. International Neural Network Conference1990”, pages 305-308, Dordrecht

Vesanto, Juha, Johan Himberg, Esa Alhoniemi, and Juha Parhankangas. 2000. *SOM Toolbox for Matlab 5* . Report A57, Helsinki University of Technology, Neural Networks Research Centre, Espoo, Finland.