Learning Morphology by Itself[1]

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

Ivan Herreros
Univesitat Pompeu Fabra, Barcelona, Spain
ivan.herreros@ilc.cnr.it

Abstract

The paper reports on a few experimental results of a computer simulation of learning the verb morphology of Italian, English and Arabic with the same type of neural architecture based on Kohonen's self-organizing maps. Issues of the mental organization of the resulting morphological lexica are explored in some detail and discussed in the light of the differential distribution of regular and irregular inflections in the three languages. It is shown that typologically diverse, non trivial aspects of the underlying paradigmatic structure of the three verb systems effectively emerge through sheer exposure to realistic distributions of verb forms devoid of morpho-syntactic content. We argue that these results go a long way towards explaining how global organization effects in the mental morphological lexicon may eventually result from local word processing steps.

1.1 Introduction

The developmental acquisition of the inflectional system of a language requires the fundamental ability to identify, on the basis of a child's exposure to its unanalysed parental input, a repertoire of formal means of marking morphological contrast. In a deliberately simplified version of this task, the child's input can be assumed to be an unstructured list of independent word forms, already properly segmented out of their embedding phonetic stream, and perceived by a learner according to a certain probability distribution. Although this helps to focus on issues of word internal structure only, the task of morphological marker identification remains, for a number of reasons, a considerably hard one.

First, morphological markers are known to wildly vary cross-linguistically (Bybee 1985, Anderson 1992, Croft 2001, Stump 2001, Haspelmath 2002), thus leaving the learning child with an exceedingly unconstrained space of alternative hypotheses for word segmentation, ranging from affixation to templatic structures and reduplication. Secondly, they are poorly salient from a perceptual viewpoint, as they tend to appear in

---

[1] The present paper is the outcome of a joint, highly cooperative effort. However, for the specific concerns of the Italian and Spanish Academies, Ivan Herreros is responsible for sections 3, 4 and 5 and Vito Pirrelli for sections 1, 2 and 6.

G. Booij, *et al.* (eds.), *On-line Proceedings of the Fifth Mediterranean Morphology Meeting (MMM5)*
Fréjus 15-18 September 2005, University of Bologna, 2007.
URL http://mmm.lingue.unibo.it/

phonologically weak, often unstressed, word boundary positions. Moreover, they convey fairly abstract and procedural semantic content (*i.e.* morpho-syntactic properties), having very few if any perceptual correlates in the grounding environment where words are uttered. Finally, when a language offers more than one realization of a given array of morpho-syntactic properties (indeed an unmarked case in the entire Indo-European family), multiple markers appear to cluster in paradigmatically-related classes, whose identification is part and parcel of the process of mastering the selection of the proper inflectional material, given a word's inflectional class. All in all, learning the inflection system of a language requires the development of a highly abstract classification system (which we may dub, in traditional linguistic terms, a "grammar") that, far from being an epiphenomenal by-product of a basically unstructured, whole-word lexicon, plays an active role in both on-line word processing and lexical access and representation (Marslen-Wilson et al. 1994).

In many respects, the learning task is reminiscent of the Harrisian goal of developing linguistic analyses (and ultimately a linguistic ontology of basic categories and atomic constituents) on the basis of purely formal, algorithmic manipulations, traditionally known as *discovery procedures*, of relatively raw language data (Harris 1951). Here (as with the problem of learning inflection broached above) a level of linguistic explanation is attained by first developing a *generalization method*, to then assess the obtained generalizations against some established theoretical background.

However, as we shall see in more detail in the following section, morphology learning cannot be simply equated with the linguist's job of establishing an ontology of morphological markers. Linguists rely on an extensive battery of *a priori* procedural knowledge (such as "morphologically complex words can be segmented exhaustively into non-overlapping constituent morphemes", "allomorphs tend to be arranged into a minimum number of disjunctive paradigm-based classes" etc.). This knowledge plays a fundamental role in ensuring convergence of Harrisian procedures on the sort of empirical generalizations aimed at by linguists. We are thus faced with the issue of whether children grow up equipped with the same battery of knowledge biases. In other words: where does all these *a priori* assumptions on word structure come to a learner from? Can we identify some basic cognitive mechanisms that are primary and foundational in the ontogenetic development of language acquisition with respect to more elaborated and specific categories of linguistic knowledge?

To address these questions, this paper presents a computer model of morphology learning that is intended to portray the learning task of marker identification as a process of *emergence* of morphological structure in the learner's mental lexicon. The approach is aimed at addressing a number of well-known aspects of cognitive development, such as the role of fluency and entrenchment in the ontogenetic development of procedural knowledge (Anderson 1993, Boyland 1996), the impact of sequential distributions on aspects of reduction in the individual articulatory gestures of word production (Bybee 2002), morphological irregularization, global effects in the morphological organization of both lexicon access and lexical representation (as opposed to whole-word models of the speaker's mental lexicon), well-known effects of local similarity in on-line morphological processing (Albright 2002), graded morphological structure (Hay and Baayen 2005). To anticipate some of the conclusions we shall draw in the final part of the paper, computer modelling of language learning, with its strong reliance on probability distributions and machine learning algorithms, may apparently bear little resemblance to traditional theoretical accounts of inflectional morphology. It may turn out that the fine-grained levels of explanation offered by computational simulations of morphology learning are not straightforwardly amenable to traditional grammatical categories. Yet, we agree with Goldsmith and O'Brien (2006) that simulating the emergence of complex levels of morphological organization in the mental lexicon is by no means incompatible with the view that speakers internalize a complex body of abstract linguistic competence. As we will show in the following pages, such a body of abstract knowledge is more intimately related to usage-based aspects of the language input than some theoretical linguists are ready to recognize.

The paper is structured as follows. Section 2 overviews different approaches to the problem of morphology learning in the light of the above-mentioned cognitive requirements and recapitulates some of the hardest challenges in modelling what we know about human morphological behaviour. Section 3 provides an introductory description of so-called Self-organizing Maps (Kohonen 2001), a member of the family of competitive neural networks exhibiting a topological behaviour that is particularly suitable for modelling the dynamics of lexical organization and on-line morphological processing.. Sections 4 e 5 outline the neural architecture used for our experiments and review a few learning results obtained on typologically diverse training data. Finally, in section 6 we draw some conclusions, carve out our future research agenda and sketch some prospective work.

2. Background

The acquisition and mastering of productive systems of inflectional morphology in natural languages are known to be extremely difficult tasks. Most adult second language learners develop relatively fixed syntactic constructions, with words typically occurring in one morphological form only (Klein and Perdue 1997, Wilson 2003). Similarly, pidgin and Creole languages are characterised by a relatively impoverished system of inflectional morphology. Moreover, inflectional competence, in both adults and children's language behaviour, tends to be relatively brittle and break down fairly quickly under various kinds of processing pressure and language impairments (Dick et al. 2001).

In this section, we shall focus on what we take to be a logically preliminary step in morphology learning: the process of scanning a word form through, to search for its morphological formatives. In particular, we shall mainly be concerned with the issue of identifying markers of inflectional categories such as person, number, gender, tense and mood, which are known to form the grammatical backbone of conjugational paradigms and constitute a primary goal of early efforts of morphology learning in child language maturation. The problem is traditionally conceptualized as the task of splitting an inflected word form into its constituent morphemes. As the notion of morpheme in both theoretical and cognitive linguistics has been the locus of much controversy over the last thirty years or so (a debate upon which dust does not seem to have settled yet), we deliberately sidestep the issue of the ontological status of inflectional and lexical formatives, to portray the task under scrutiny as the somewhat preliminary and more fundamental goal of identifying recurrent elements of formal realization of morphological contrast. To be more concrete, we would like to focus on the ontogenetic process by which an Italian child, exposed to verb forms such as *amiamo* 'we love', *canto* 'I sing', *vengo* 'I come' etc., is able to identify the recurrent segmental units

'*-iamo*' and '*-o*' as typical (albeit not necessarily exclusive or minimal) carriers of information about person and number in the Italian present indicative sub-paradigm.

Even when stated in these simplified terms, the problem is considerably harder than expected. First, the child has no way to know, a priori, where inflectional formatives should be looked for within a verb form. Her/his search space is thus potentially very large: a huge haystack with comparatively few morphological needles. Secondly, the amount of formal redundancy exhibited by verb forms in a given language goes well beyond the limited range of recurrent morphologically relevant markers. Rhyming words, false friends, false prefixes and the like are virtually ubiquitous and tend, at least in principle, to obscure morphologically relevant analogies. We may refer to this as the background noise problem. On top of that, relevant analogies happen to be often confined to one segment only, in the perceptually weak coda of a word final syllable. Even in the same language, prefixation, suffixation and stem alternation often present themselves simultaneously in tricky combinations. Particularly in highly frequent irregular or subregular verb forms, more strategies of morphological marking appear to often be overlaid, to the point that formal discontinuity is a prominent feature even of those languages that do not exhibit non-concatenative morphology.

It is very difficult, for a non linguist, to disentangle herself/himself from such an intricate coil of input evidence. The machine learning literature, with its large array of assumptions about algorithmic searching of formal redundancies, has enormously contributed to shed light on these and related issues (Pirrelli 2003 and references therein). The apparently naïve question at the heart of our investigation is thus the following: what does it take for a child to become sensitive to few morphologically relevant formal analogies and remain blind to the very many ones bearing no or scanty relationship to abstract principles of grammatical organization in the morphological lexicon? Linguists have often confronted themselves, in either direct or indirect ways, with this puzzling question. In the following section we briefly recapitulate some of the most influential answers in the literature. This will lead us to talk about the problem of the possible sources of the knowledge required by a child to home in on the appropriate battery of language-specific markers of inflectional features.

2.1. Nativism

According to a well-established nativist position, emanating from the generative approach to adult grammar competence of Chomskyan inspiration, children are equipped with an innate set of options for acquiring distinct language types. According to this view, the extensive cross-linguistic variation exhibited by the morphology of human languages can be explained by positing language-specific ways of setting these options, called "parameters", in the grammar word module. In particular, it has been argued that the child scans her/his linguistic environment for designated structures or "cues", to be found in the mental representations which result from hearing, understanding and parsing words (Lightfoot 1999, Hyams 1986, 1996). Cues which are realized only in certain typological families of grammar constitute the parameters. For example, upon understanding a word form like *sneezing*, the child comes up with an abstract representation such as $[_{V}[\textit{sneeze}]_{progr}[\textit{ing}]]$ allowing her/him to set a word-final parameter concerning the position of inflectional markers in the English verb. Surely, the parameter can only be set when a child has already homed in on a partial analysis which treats *sneeze* and *-ing* as separate sub-word constituents, the latter being interpreted as a marker of abstract morpho-syntactic features. Thus, the availability of valuable cues for morphology learning *presupposes* an appropriate segmentation of *sneezing* rather than providing a principled solution to the haystack search problem.

2.2. Connectionism

Over the last twenty years, connectionism has challenged the symbolic view of morphological processing dominant in the Chomskyan tradition to provide a coherent alternative approach to the issue of learning word internal constituents. One of the most articulated and full-fledged recent illustrations of this proposal (Plaut and Gonnerman, 2000) views morphology as an interface realm, emerging as a pattern of activations in the layer of hidden units mediating the relationship between lexico-semantic and phonological word representations in an artificial neural network (Figure 1).

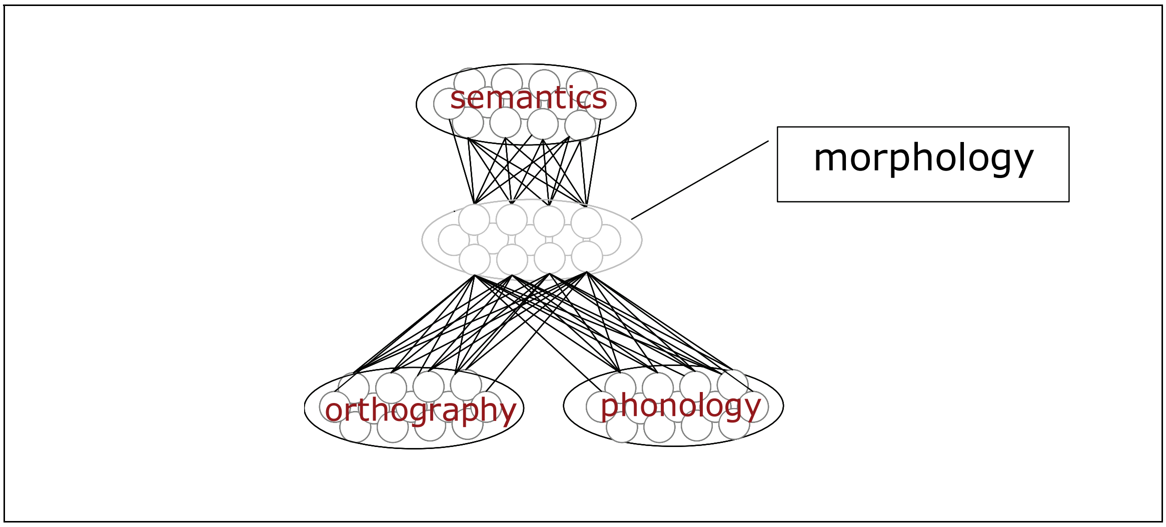

Figure 1: A connectionist framework for lexical processing (adapted from Plaut & Gonnerman, 2000)

To the extent that a particular surface pattern occurs in many words and maps consistently to certain aspects of lexical meaning, the representation conveyed through the internal (hidden) layer as an array of activation states will come to reflect this mapping, and will process it relatively independently of other parts of the word. This developmental process accounts for gradient effects of morphological structure, with intermediate degrees of morphological transparency being related to intermediate degrees of either phonological or semantic transparency (Plaut and Gonnerman 2000, Hay and Baayen 2005). For our present concerns, the interest of this proposal rests on the possibility that the child's hypothesis space be effectively constrained by relating the search for formal redundancies to the existence of shared semantic representations. This should considerably limit the combinatorial explosion of useless mappings between deceptively similar word forms (background noise), but does not seem to address our segmentation problem in a principled way, for two basic reasons. First it requires that children have access to highly complex, fully developed lexico-semantic representations, whose early availability in the parental input to the child is somewhat moot. In fact, we have evidence that the acquisition of abstract morpho-syntactic categories and a full understanding of their role in language processing tend to occur at a comparatively late stage of language maturation, when the child has already mastered those aspects of morphological realization and marker selection we are presently concerned with (Clahsen 1989, Wilson 2003). Secondly, connectionist representations of the phonological input of inflected word forms do not offer a principled account of the word mapping problem. This point is illustrated by the input word representations used by Plunkett and Juola (1999) for experiments on learning English noun plural inflection (illustrated in Figure 2 for the words *cats* and *oxen*). Input representations are obtained by integrating phonological and morphological information into a fixed-size template-like structure (where the segment sequence /ts/ in /kAts/, for example, is, contra phonological evidence, split by an intervening empty vowel slot), with the result of enforcing a built-in alignment between input representations of words selecting different inflectional endings. The alignment has the effect of slipping in a strong language-specific bias that appears to *presuppose,* rather than explain, the problem we are presently concerned with.

C	C	C	V	V	C	C	C	V	C
#	#	k	#	A	#	#	t	#	S
#	#	#	#	O	#	k	s	E	N

Figure 2: A templatic input representation for English noun plurals (from Plunkett and Juola 1999)

2.3. Distributionalism

Harris' assumption that morphological categories can be derived mechanically from an analysis of the distributional properties of word forms in context has the potential of addressing the range of questions we are concerned with in this paper. According to Harris' view (Harris 1951), identification of relevant inflectional formatives is the final result of building a statistical model of the way overt, perceptually salient strings of phonological segments follow each other in the language input which is the ultimate object of linguistic investigation. Endorsing a somewhat radical mistrust in the role of semantic or more generally non perceptually overt knowledge in language analysis (Matthews 1993), Harris delineates a purely formal methodology whereby the only evidence available to the linguist is made up out of strings of linguistic units and their distributions. His approach, after a long-lasting obsolescence, has recently played an inspirational role for a number of machine learning approaches to unsupervised morphology acquisition (Gaussier 1999, Goldsmith 2001, Schone and Jurafsky 2000, Creutz and Lagus 2004, Wicentowsky 2004).

In a recent adaptation of Harris' ideas, John Goldsmith (2001, 2006) casts the distributional hypothesis into a powerful information theoretic framework, known as Minimum Description Length (MDL, Rissanen, 1989). Starting form the assumption that morphological information about a language can hardly be reduced to *local* information about letter bigrams or trigrams of that language, Goldsmith frames the task as a data compression problem: "find the battery of inflectional markers forming the shortest grammar that best fits training evidence", where i) a grammar is a set of paradigms defined as lists of inflectional markers applying to specific verb classes and

ii) the training evidence is a text corpus. The task is a top-down global optimization problem and boils down to a grammar evaluation procedure. Given a set of candidate markers, their probability distribution in a corpus and their partitioning into paradigms, MDL allows calculation of i) the length of the grammar (in terms of number and size of its paradigms) and ii) the length of the corpus generated by the grammar (*i.e.* the set of inflected forms licensed by the grammar according to a specific probability distribution). In MDL, the notion of length is derivative of the information theoretic notion of the number of bits required to encode linguistic units, whether they are stems, suffixes or word tokens. Intuitively, minimising the length of the corpus in bits requires that very frequent tokens should be assigned a shorter bit code than less frequent tokens. Minimising the length of the grammar, on the other hand, requires that frequently used paradigms are given preference to rarely used ones, as the cost of encoding a rare paradigm in bits is very high. Hence, a good language model is the one where the sum of the length of the grammar and the length of the corpus generated according to the probability assigned by the grammar is smallest. This policy disfavours two descriptively undesirable extremes: a corpus-photograph model, with a very long grammar where each verb form has, as it were, a paradigm of its own, such that the inflected forms generated by the grammar have the same probability distribution found in the corpus; and a very short but profligate model, with one paradigm only, where any verb combines with any marker according to the product of their independent probability distributions, thus generating many word forms that are not attested in the training corpus (including *goed* for *went*, *stricked* for *struck*, *bes* for *is* etc.).

From a cognitive perspective, Goldsmith's approach has the merit of addressing the problem of morphology learning with no recourse to prior language-particular knowledge. Furthermore, he adopts a mathematical framework where the development of morphological knowledge can be viewed as the emergent result of data compression, arising, both phylogenetically and ontogenetically, from the pressure of keeping a potentially unbounded amount of lexical knowledge in a finite memory store. We find these ideas fundamentally correct. On a less positive note, in Goldsmith's approach the issue of morpheme segmentation is kept separate from that of morpheme inventory evaluation, both logically and algorithmically. The two learning phases make no contact, so that we are left with no principled answer to the problem of the interplay between word processing and morphological organization in the speaker's mental lexicon: does morphological organization play any role in word processing?

Moreover, it is hard to see how a child learning morphology can possibly be engaged in such a top-down search for global minima. What we know about word processing in human subjects supports the view that speakers are extremely sensitive to *local* similarity maxima and tend to analyse and generate novel word forms predominantly (if not exclusively) by analogy to their closest cognates. For example, Italian speakers appear to be able to use fine-grained classes of verb stems to assign them the appropriate conjugation paradigm. According to Albright (2002), Italian speakers are able to assign a 0.937 conditional probability to the event that an *X*[end] verb stem is inflected for the second conjugation class. This means that when an Italian speaker is exposed to the nonce 1s present indicative form *trendo*, (s)he is almost certain that its infinitive is *trendere* (and not *trendare* or *trendire*). Most evidence for such an acute sensitivity to local similarity comes from *irregularly inflected* verb forms (but see again Albright 2002, for similar effects with regular verbs) and is often contrasted, in the psycholinguistic literature, with the somewhat opposite tendency towards using default rules in the production of *regularly inflected* forms (Say and Clahsen 2001). According to many scholars (Pinker and Prince 1988, Prasada and Pinker 1993, Marcus et al. 1995 among others), the contrast supports a *dual route model* of word processing and learning: irregular forms are stored in full and are generalized over by local similarity, while regular forms are stored and indexed by their roots and affixes and produced by default rules of some kind. Other scholars oppose to such view and argue in favour of a unitary underlying mechanism accounting for both regular and irregular forms (Rumelhart and McClelland 1986, Plunkett and Marchman 1991, Bybee 1995, Ellis and Schmidt 1998 among others).

We have no room here to address this debate in any detail. Suffice it to point out that, for our present purposes, we are faced with an apparent paradox. We agree with Goldsmith that learning the morphology of a language can be framed, in machine learning terms, as a global optimization problem: morphologically relevant analogies (unlike local, potentially misleading similarities) emerge from a global analysis of the available input evidence. On the other hand, we are forced to reconcile this truth with the fact that speakers use local analogy-based strategies to develop morphological generalizations. The somewhat paradoxical question then is: how can a learner home in on global, paradigm-based analogies on the basis of local processing strategies?

In the remainder of this paper we intend to show that Self-Organizing Maps (SOMs), a particular family of artificial neural networks, can offer an interesting way out of this apparent paradox. As we shall see, SOMs can develop topological maps of input stimuli where the latter are organized according to global classification criteria. This is so in spite of the fact that SOMs learn and process input stimuli on the basis of principles of purely local analogy as will be shown in the following section.

3. Self-Organising Maps

3.1. Brain Maps

A Self-Organizing Map (hereafter *SOM*, Kohonen 2001) is an unsupervised machine learning algorithm drawing considerable neuro-physiological inspiration from the behaviour of so-called “brain maps”. Brain maps are medium to small aggregations of neurons found on the cortical area of the brain that are involved in the specialized processing of specific classes of sensory data. Processing simply consists in the activation (“firing” in neurophysiological terms) of a certain neuron (or neuron aggregation) each time a particular stimulus is presented. The associative links between a stimulus and its firing neurons are described, in neurophysiological terms, as “mapping”. A crucial feature of the sort of mapping performed by brain maps is that similar stimuli fire nearby neurons. As we shall see in more detail later on, such a local sensitivity to similarity in the presented stimuli develops inside a globally ordered topological structure. This is so because local mapping must obtain over the entire brain map area, thus enforcing an incremental principle of global organization of firing neurons. Examples of brain maps are i) the *somatotopic* map where stimuli generated in different parts of the body are mapped onto different specialised areas, ii) the *tonotopic* map where neurons respond to sound stimuli according to the frequency of the sound or, iii) the colour map on the visual area V4. The genesis of such brain maps is also interesting for our present cognitive concerns. Although some of them can be considered as genetically pre-programmed, there is evidence that at least some aspects of such global neural organizations emerge according to the sensory experience of the subject (Jenkins et al. 1984, Kaas et al. 1983).

In 1984, Teuvo Kohonen described an iterative, unsupervised Artificial Neural Network (ANN) exhibiting some salient characteristics and behaviour of brain maps. Each unit/node of an ANN can be viewed as a receptor neuron that reacts to (or is activated by) a particular class of stimuli only. A node is an independent processing unit associated with a small memory trace that stores the stimulus the node is sensitive to. The more faithful the trace, the more sensitive the receptor. From this perspective, simulating the behaviour of a brain map is tantamount to developing an incremental ANN where similar stimuli trigger topologically neighbouring nodes.

The SOM learning algorithm is iterative. At each iteration the network is exposed to a random input stimulus. The first phase of the iteration consists in a network activation, culminating in the identification of the best matching unit on the map. The best matching unit (*BMU*) is the receptor whose memory trace happens to be closest to the current input. Memory traces are sometimes called prototype vectors (because they are represented as vectors), but they can also be referred to simply as the unit memory content. If we consider that the input is just another vector, the search for the best matching unit simply consists in finding the map node that contains the vector most similar to the input. Returning to the analogy to brain maps, the best matching unit in an ANN plays the role of a real neuron(s) being fired in a brain map.

In a SOM, the activation part of a learning iteration is followed by an updating phase. The memory contents of a certain number of map units are updated for them to look closer to the new information provided by the last input stimulus. An update consists in adjusting a number of memory traces to the input pattern just presented to the map. Using a slightly far-fetched metaphor, we can describe the neuron memory as a camera film being repeatedly exposed to an image at very short time intervals (learning iterations) for an amount of time insufficient for a clear image to imprint the film one-shot. At each exposure, the image on the film resembles more and more closely the input image the film is exposed to. It is important to appreciate that the update process is undergone neither by all neurons, neither at the same rate for all involved neurons. In fact two parameters, the neighbourhood radius and the learning rate, govern the learning process in determining, respectively, the number of units being updated at each iteration and the amount of incremental adjustment at each time tick. Both parameters play a crucial role in the dynamics of the learning process and decrease as learning progresses.

3.2. The neighbourhood radius

After the BMU is identified, a number of neurons are updated: these include the BMU itself and a set of its neighbouring units on the map, within a distance (from the BMU) defined by the neighbourhood radius. At the beginning of the learning process, the radius is long enough to guarantee that large neighbouring areas of the map are updated at each iteration. This ensures that a global order is enforced upon memory traces. Finer-grained relationships are learned at later stages, when the neighbourhood radius is progressively reduced in the course of learning. This defines a fundamental dynamics of a SOM learning trajectory to which we shall return later.

3.3. *The learning rate*

The learning rate defines the amount by which the memory content of each unit is modified at each iteration. At early stages of learning, the rate is kept high, thus allowing memory traces to quickly adjust to input data. As learning progresses, however, the rate decreases and memory traces gain in stability.

3.4. *Plasticity*

The joint effect of the dynamics of both neighbourhood radius and learning rate defines the so-called network plasticity, i.e. the capability of a map to modify its content to adapt it to input data. At early stages, the map content is extremely unstable and adaptive due to a long neighbourhood radius and a high learning rate. In the end, the map plasticity reduces considerably, thus allowing for a process of fine tuning only. As a result of this joint dynamics, a SOM can learn the global order underlying input data only at early stages, when plasticity is high and the map topology can be modified easily. By the same token, it is only when plasticity goes down and the network becomes more stable that fine-grained distinctions are acquired.

3.5. *Frequency effects on a self-organising map.*

SOMs are very sensitive to input frequency. To better understand this point, it is important to bear in mind that the basic task of a SOM is to accommodate input stimuli on its surface by associating them with corresponding memory traces. If there is enough room on the map, then every input stimulus will be assigned a faithful trace though learning. For lack of room on the map, on the other hand, similar input stimuli will tend to compete for the same memory traces. In this competition, both stimulus token frequency and stimulus class frequency play a key role. By their being repeatedly exposed to the map, high token frequency inputs are bound to carve out a map area where they are memorized faithfully, even if they form a class of their own: due to their high token frequency, they can in fact win the competition by themselves. On the other hand, low token frequency stimuli will leave a memory trace on the final map only if they are part of a high frequency class of stimuli, that is a class where the sum of token frequencies of its member is high.

This has a simple probabilistic interpretation. For example, in the case of the Italian past participle, if we consider the class of verb forms ending in -*ato*, its class frequency will tell us how likely we are to find a member of that class in a given corpus. This has also implications in terms of memory traces. When memory traces are exposed not to a single stimulus type, but to an entire class of similar stimuli, they will tend to reflect, for lack of sufficient room, what the class members have in common. Let us suppose we are running a toy experiment where the word *cantato* (‘sung’) leaves a memory trace on the final SOM only because it is part of a high frequency class whose other two members are *amato* (‘loved’) and *pensato* (‘thought’). As these three forms share the same memory trace, the latter will reflect the commonalities partaken by them, for example the ending -*ato*. The case illustrates a simple effect of “generalization as a shortage of memory”.

On the other hand, if a high frequency stimulus forms a class of its own, the particular memory trace (or memory area) fired by it on the map should be able to represent it faithfully. This is what happens, for example, when we find that the form *said* is fully memorized on an English past participle map. Note that this situation is the mirror image of what we found out in the previous paragraph: in fact, full storage of a very frequent input leaves no room for generalization. The natural question at this juncture is: if words with high token frequency are fully memorized, what is the relationship between them and other partially memorized, less frequent words which nonetheless belong to the same class as the former?

So far we have discussed how both token and class frequency affect a) the possibility for a stimulus to be learned by a SOM and b) the kind of memory trace the stimulus is likely to leave on the output map. In both cases we have been looking at the end result of learning. It is now time that we turn to discussing in some detail what happens *during* the learning process as such.

In the process of learning, traces memorised on a SOM slowly approximate original input representations. If an input is presented a number of times exceeding a certain threshold, the SOM will contain a memory trace with a faithful representation for that input. According to the dynamics of the learning process, however, the minimum threshold goes down as learning progresses (and the neighbourhood radius decreases). In other words, the rate at which an input is presented to the map has a direct impact on the rate at which the input is “learned”. Intuitively, single tokens are going to be learned at a rate which is proportional to their frequency. Similarly, frequent stimulus classes are learned earlier than less frequent classes (if the latter are going to be learned at all).

To be more concrete let us turn to the map of Italian past participles. As *stato* (‘been’) is a very frequent item in the training corpus, we expect the following three consequences: a) *stato* will be memorized on the map; b) it is likely to have one or more dedicated firing neurons, that is neurons with a faithful image of the input; c) the image of *stato* will appear at an early stage of the learning process. What about less frequent forms like *amato* or *cantato*?

In a SOM, the global order of the map is fixed at early learning stages, at around the same time frequent forms are memorized in full. In fact, it turns out that high-frequency items play a crucial role in i) shaping the (high-level) topography of the resulting SOM, and ii) conditioning the topological distribution of the remaining information. In fact, very frequent input items act as prototypical representatives for their whole class over the first learning stages, thus anchoring their class representation throughout the learning process. It is only at later stages, when fine-grained information about the remaining members of the class is separated from the information about the class prototype that the whole class is fully learned. The complex paradigmatic structure of Italian past participles is particularly suitable to illustrate this kind of behaviour.

4. Experiment Architecture

The computational architecture we developed for these experiments consists of two hierarchically connected SOMs (Fig. 3), whose mode of interaction is reminiscent of Time Delay Neural Networks (Waibel et al., 1989). Input word forms are strings of alphabetic characters. At each exposure, an input verb form triggers a series of activations on the first level map. All activations triggered by the same input form are integrated into an "activation image" that is in turn processed by the second level SOM. This way, the first level image plays the role of a short term memory temporarily storing the character-based information relative to a single input form, for this information to be processed wholesale by the second level map.

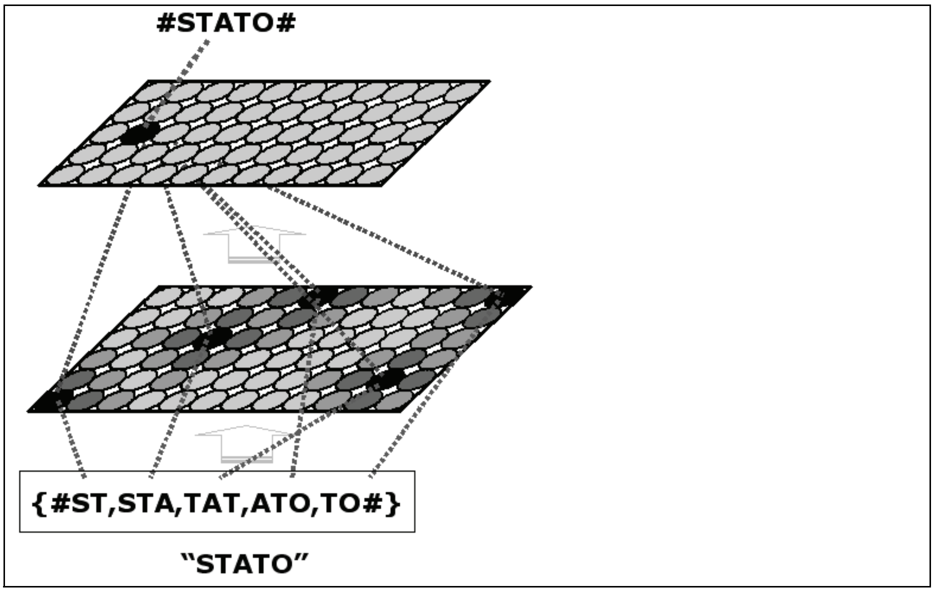

Figure 3: A two-level SOM architecture

Activations on the first level SOM are produced by consecutive time-bound scans of the input form. Each activation represents the map's response to a sub-context of the whole input, whose fixed size is measured as the number of alphabetic characters it contains. For the current set of experiments, the context size was set to 3. The second level SOM takes as input the output image of the current input form on the first level map. The upper level SOM clusters word forms according to their activations images on the lower level SOM. Another way to look at the first level SOM is as a perceptual interface between the raw character-based representation of an input form and receptors on the second level map. The fundamental benefit of this two-staged processing strategy is that the dimensionality of each "activation image" remains fixed, independently of input size.

5. Experimental Results

In this section, we report the experimental results of a few computer simulations of learning verb forms in Italian, English and Arabic. In the Italian and English experiments, we contrast the differential results obtained by training the map on two data sets for each language, one where verb forms are presented to the map according to their frequency distribution in a corpus, the other one where training data are distributed uniformly. The comparison sheds light on the role of frequency in the learning dynamics by exploring two considerably different verb systems (Italian and English) in terms of the relative prominence of sub-regular verb forms with respect to fully regular ones. On the other hand, the Arabic results are based on one training configuration only, whereby fully vocalized verb forms are presented to the map according to their frequency distributions in the LDC Arabic corpus (Maamouri et al. 2004). For each experiment, we focused on paradigmatically homologous verb forms: past participle masculine singular forms for Italian, past tense forms for English, perfective masculine third singular forms for Arabic.

The experiments are intended to cover a fairly wide range of both typological and structural dimensions of cross-linguistic morphological variation. English and Italian verb forms are contrasted with Arabic data along the concatenative vs non-concatenative dimension of inflectional marking. From this standpoint, the crucial issue is whether it is possible for a single map to simulate the differential processes of acquisition of typologically as diverse morphological constructs as inflectional endings, continuous stems, word patterns and discontinuous stems on the basis of uniform requirements on input representation. This is not trivial, since, as we saw above, the representation requirements for English and Arabic data are potentially conflicting and generate hard alignment problems.

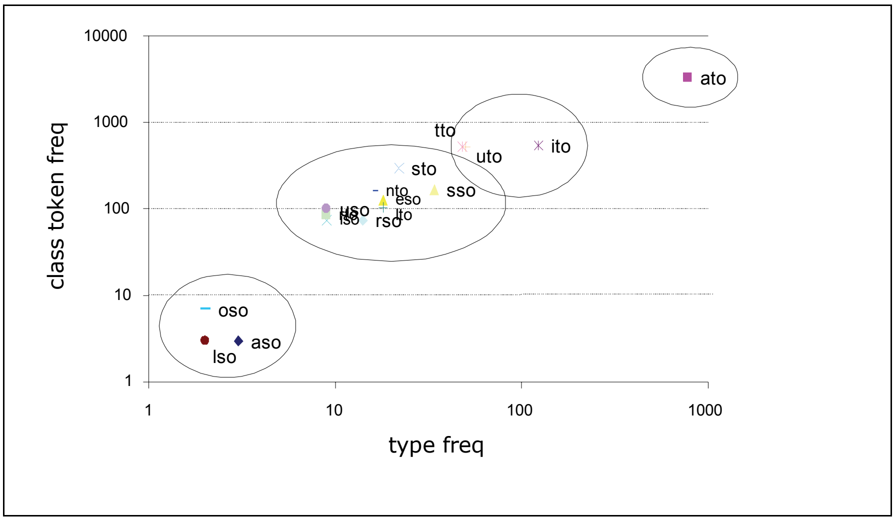

Figure 4: Type vs. class token frequency in the Italian past participle

Another important dimension of variation this set of experiments is intended to shed light on concerns the different distribution of regular and irregular inflections in languages such as English and Italian and the way such differences may impact morphology learning. Figures 4 and 5 illustrate this point in connection with the distribution of past participle and past tense forms in the two languages, by type and class token frequency plotted on a log scale. Forms are grouped according to loosely defined inflectional classes, each including inflected forms sharing the orthographic rendering of the word final syllable nucleus. Admittedly, the criterion is fairly crude and fails to cluster together forms such as *left* and *felt* whose past tense formation processes are very similar. Nonetheless, the resulting classifications retain some morphological plausibility and stake out the space of formal variation a learner is exposed to.

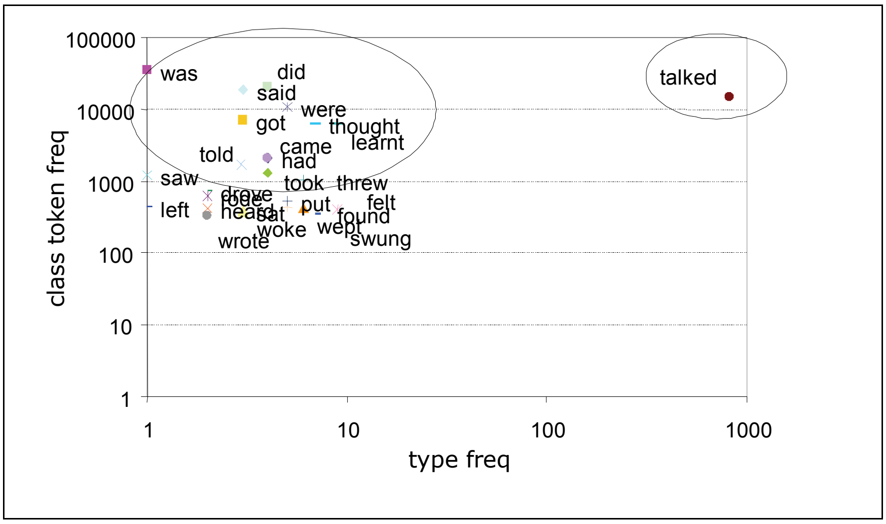

Figure 5: Type vs. class token frequency in the English past tense

The difference in distribution between Italian and English data is striking. Italian forms are scattered along a fairly uniform continuum, with subclasses of irregular forms exhibiting increasingly prominent gang effects in terms of their type cardinality as we move from left to right along the x-axis. On the other hand, English forms can sharply be divided into two groups: irregular forms on the left-hand side of our plot, forming a constellation of scantly represented morphological sub-classes (whose cardinality hardly exceeds the 10 units) and the class of regularly inflected forms on the other hand, covering the vast majority of English verb types. Besides, the Italian distribution shows a prominent log-linear correlation between type frequency and class token frequency, totally missing in the English past tense. This fact, as we shall see in a moment, has significant consequences on the learning dynamics of the two systems.

5.1. Italian past participles

Input forms consist of 470 different singular past participle forms (for a total amount of 5157 tokens) from the Italian Treebank (Montemagni et al. 2003). The highest frequent form is *stato* (‘been’, with 382 occurrences), followed by *fatto* (‘done’, 180), *detto* (‘said’, 131), *visto* (‘seen’, 77) and *avuto* (‘had’, 70). The least frequent forms appear only 3 times in the Treebank and cover 106 different forms, 75 of which ending in *-ato* (first conjugation), 15 in *-ito* (third conjugation) and only 2 in *-uto* (second conjugation). Of the remaining 14 form types of frequency 3, all undergoing a sub-regular past tense formation (Pirrelli, 2000), only two are verb base forms (namely *sciolto* and *stretto*) while the remaining 11 are derivatives such as *esteso*, *rimosso* and *trascorso*. As the training corpus is a collection of newspapers articles, speech report verbs such as *dichiarato* (‘declared’), *aggiunto* (‘added’) or *annunciato* (‘announced’) are among the most frequent form types. Surely, these figures prevent us from taking this experiment representative of the typical input evidence an Italian child is exposed to in the course of her/his morphology maturation. Nonetheless, our word distributions, however not as realistic as we would like them to be, do reflect, to a certain degree of approximation, a general bias towards consistently sub-regular high-frequency Italian verb forms such as *detto* ('said'), *fatto* ('done'), *visto* ('seen'), *chiesto* ('asked') etc., that happens to hold independently of variation of topic, gender and pragmatic grounding.

We simulated two different learning sessions: one where verb forms are presented to the map according to their frequency distribution in the Italian Treebank, the other where training data are assumed to be distributed uniformly. Figure 6 gives two snapshots of the state of a second level map in the two learning sessions at the same (early) stage. Grey triangles highlight map nodes that are fired when past participles of the *-sto* class (e.g. *visto* 'seen' and *chiesto* 'asked') are presented to the map. Black triangles highlight nodes that are sensitive to the *-tto* family (*fatto* 'done', *detto* 'said' etc.).

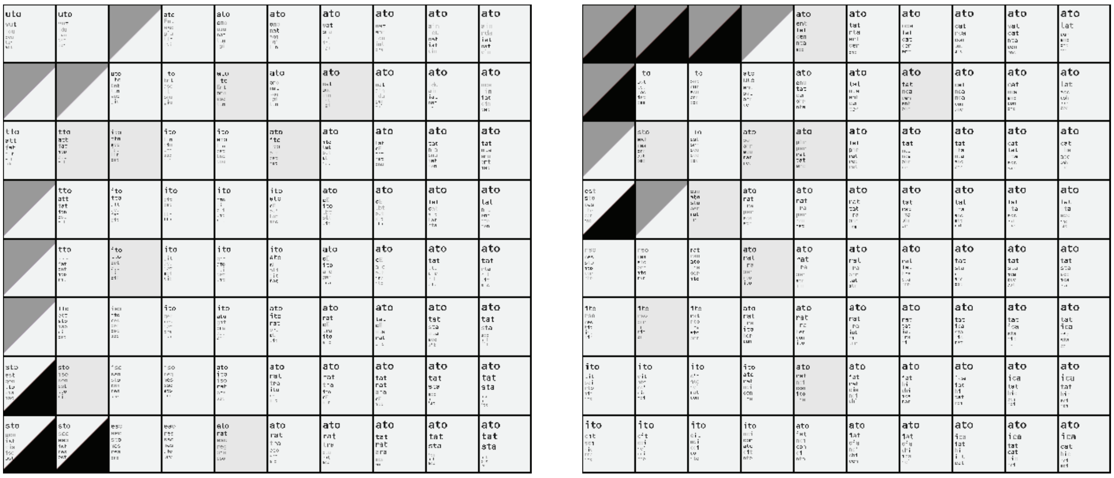

Figure 6: The Italian past participle

The important difference between the two snapshots is that the map trained on token frequencies neatly separates the two verb classes, while the other map tends to confound them: in other words, the former map develops entrenched, differential specialisation for *–tto* and *–sto* ending forms quite early on the basis of their token frequencies, while the latter SOM more reluctantly converges towards specialisation, for lack of evidence on token distribution. Later in the paper, we explain this developmental difference by arguing that very frequent forms like *visto*, *fatto* and *detto* tend to act as *prototypes* of their own class.

5.2. English past tenses

The experiment input consists of the 548 most frequent past tense forms in the British National Corpus (Leech 1992). The top-most ranked such forms are *was* (34836 occurrences), *did* (20247), *said* (18051), *were* (10570) and *had* (9573), accounting for a total of 93278 occurrences, out of all 141501 past tense forms attested in the training corpus. Like with the Italian experiment, we simulated two learning sessions, with and without token frequencies. Interim results of the two sessions, at comparable learning stages, are depicted in Figure 7, showing two dramatically different topological structures. In the right-hand map of Figure 7, grey squares mark map neurons activated by *–ed* ending forms, while black squares are fired by *was*, *had* and *did*. In a 6x8 map grid, regular forms spread over 41 nodes, leaving only seven nodes to all remaining irregular forms. The result sets the stage for massive regularization of sub-regular forms, which are swamped by their *–ed* competitors, and seemingly lends support to the view that the two sets of regular and sub-regular past tense forms cannot possibly be learned through the same mechanism. On the other hand, the left-hand map of Figure 7 shows the results of learning by token frequencies: regular *–ed* ending forms now take up only six nodes on the map, while *was*, *had* and *did* have each a dedicated neuron. More room is left for memorizing other irregular forms, such us *said*, *paid*, etc. The evidence is in line with the intuition that irregularly inflected forms have the chance to survive the regularizing pressure of *–ed* forms, if the former are frequent enough to carve out their own dedicated area on the map by repeatedly firing a highly specialised, although comparatively circumscribed area of map nodes.

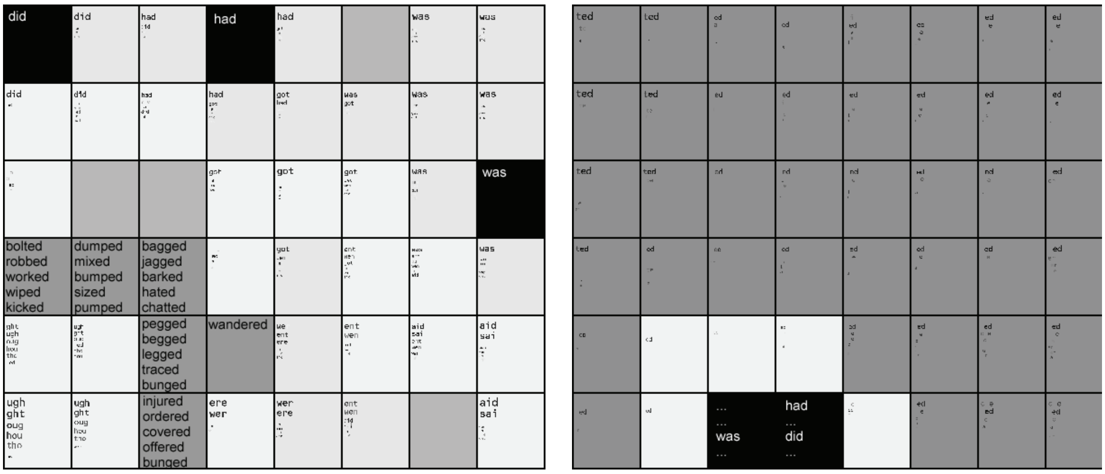

Figure 7: The English past tense

5.3 Arabic 3ps-SG perfectives

Comparable results are obtained by feeding a two-level SOM on Arabic verb forms, namely third masculine singular perfectives. This time, morphological markers do not form continuous strings of characters (as with Indo-European endings), but rather vowel patterns that are interdigitated with discontinuous roots. Since forms are presented to the map according to their corpus-based frequencies, the pattern *a_a_a*, by far the most frequent and regular one in Arabic perfective verb forms, takes over a substantial portion of the second level map, as shown in Figure 8. We have no room here to comment on the topological structure of Figure 8 in detail. Suffice it to point out at this stage that other less regular patterns of perfective formation emerge from the map, including low frequency *a_i_a* patterns. Most remarkably, high frequent forms such as *kAna* ('(s)he/it was') and *qAla* are recognised as wholes by specialized receptors (located in the top left corner of the map). Finally, it should be appreciated that the Arabic forms used for training the two-level map are given the same input representations as English and Italian forms. Nonetheless, the resulting topology consistently reflects the specific non-concatenative nature of Arabic morphology. We take this to show that our architecture exhibits a highly adaptive and convergent topological behaviour, based on a comparatively poor battery of built-in inductive biases.

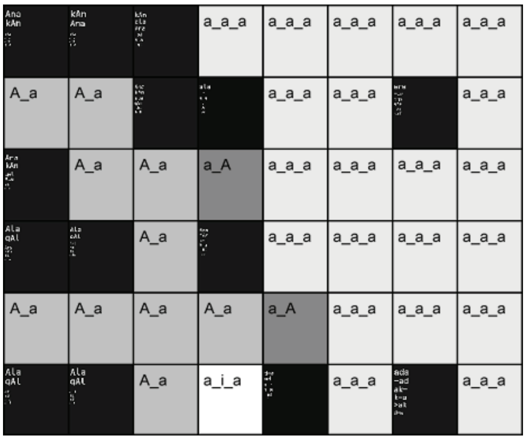

Figure 8: The Arabic perfective

6. General discussion

It is a well known fact that highly frequent forms tend to be shorter cross-linguistically, more readily accessible in the mental lexicon, independently stored as whole items (rather than being part of bigger families of paradigmatically-related forms) and thus more easily learnable and usable (Caramazza et al. 1988, Stemberger and MacWhinney 1988, Bybee 1995, Mowrey and Pagliuca 1995, Slobin 1997, Hare et al. 2001). These features make them also fairly resistant to morphological overgeneralization through time, thus establishing an interesting correlation between irregular inflected forms and frequency (Bybee 1985, 1995, Corbett et al. 2001). In the cognitive literature, it has also been shown that the existence of a type of instance that occurs with high token frequency may provide a highly relevant "cognitive anchor", serving to organise memory and reasoning about other related types (Strack and Mussweiler 1997, Goldberg et al. 2004). If we try to reconcile the latter finding with classical accounts of lexical entrenchment, we arrive at the seemingly paradoxical conclusion that irregular forms should act as "models" of the morphological organisation of the speaker's mental lexicon.

Observation of the learning behaviour of a SOM in our previous experiments can help us to understand why this paradox is only apparent. Entrenchment of a SOM memory trace is a direct function of input frequency and reflects the receptor sensitivity to input features. Similar memory traces tend to cluster in locally connected areas of the map. During training, specific, connected areas of receptors become increasingly more sensitive to specific classes of input stimuli, mimicking what we know about the functional specialization of the brain cortex. By training a SOM on a corpus-based distribution of inflected forms, then, very frequent short forms such as *is* or *did* are, at early stages of learning, the only input items to be fully memorised by receptors. These early "specialised" receptors are very likely to be fired by other, less frequent input forms (e.g. *said*, *read* or *led*) that happen to be similar to already entrenched memory traces. As lexical stems show a greater degree of formal variability than inflectional markers, memory traces of highly frequent forms tend to be fired by similarly inflected, less frequent forms. The area of the map surrounding the *did* receptor, for example, becomes more and more sensitive to *d*-ending verbs.

To sum up, we can describe the dynamic behaviour of a SOM learning the morphology of a language along the following lines:

- highly frequent forms leave deeply entrenched and highly salient memory traces that act as standards of comparison (anchors) for other similarly inflected forms
- highly frequent forms are eventually memorised in full
- less frequent forms tend to fire connected topological areas of the map that are sensitive to shared morphological markers
- the surface of each connected area is proportional to the number of form types sharing a specific marker: regular markers are thus distributed over larger areas
- formally similar markers are memorised in contiguous areas of the map, thus developing hierarchical clusters of formally graded inflections (e.g.: *-id*, *-ed*, *-t*, *-nt* for the English past tense)
- principles of SOM specialization approximate a maximally compact arrangement of memory traces.

It is important to emphasise at this stage that these promising results are obtained through unsupervised training sessions, whereby a SOM is given no indication about the possible morpho-syntactic content associated with each form. In a sense, as adumbrated in the title of this contribution, morphological classes are learned through recourse to purely morphological information only. This is interesting, as it allows us to speculate that formal principles of the morphological organization of a language can be learned by a child through sheer exposure to plain forms and their frequencies, rather than to full-fledged sign-based word representations, coupled with form and meaning. It is tempting to suggest that a child can use acquired formal principles of paradigmatic organization to regiment the proper interpretation of the morpho-syntactic content associated with inflectional endings. This suggestion is in line with the empirical evidence that children master the morphological inflection of their own language before they can use it in the appropriate morpho-syntactic contexts (Clahsen 1989, Wilson 2003).

We would like to conclude the present contribution with a few remarks. First, the developmental interplay between type and token frequency of input items throws in sharp relief the profound interconnection between entrenchment of highly frequent items and overall effects of global organization in the topology of the mental lexicon. This is not trivial and serves to reconcile two apparently contradictory but established facts in the child learning literature: i) the first stages of language learning are best described as a process of item-based rote memorization, leading to gradual development of more and more abstract morphological schemata (Tomasello 2000, in press); ii) the most frequent evidence available to a child learning the morphology of a language is in fact the most untypical and resistant to rule-based generalizations. We can explain away this paradoxical state of affairs by observing that the most frequent input items do in fact exemplify a wide range of processes of inflectional marking, thus contributing to shaping the overall organization of the child's morphological lexicon. In the early stages of learning, they act as powerful attractors of their own class mates and do so in a very focused and efficient way, since they are usually very short and reduced items, in which morphological marking has the upper hand, as it were, over lexical marking. Once the overall topology is established, the role of prototypes progressively shrinks, to eventually give way to a finer-grained organization of inflectional classes. It is at this stage, that regular patterns emerge. In our view, SOMs illustrate this dynamics in a very clear and intuitive way. Moreover they help us to gain insights into the haystack search problem we broached at the beginning of the paper. Being reduced forms, prototype attractors make it considerably easier to focus on the morphologically relevant bits of word forms.

Another related observation is that the existence of highly frequent prototypes also solves the paradoxical interplay between local processing and global, long-term memory structures. Word processing remains local throughout, but it gets progressively influenced by competition among different prototype attractors, each developing a local area of item-based influence. We suggest that the interplay of these two factors goes a long way towards explaining how global organization effects may eventually result from local processing steps.

The computational framework for morphology learning presented here leaves many issues open. At this juncture, we would like to only mention a couple of them. We have been using SOMs as topological metaphors of the mental lexicon, or, in more neuro-psychological terms, of long term memory structures. This is attractive but leaves us with the following problem: if morphological clusters develop through underspecified memory traces, how can a learner retrieve a fully inflected form? We emphasised that only highly frequent forms are memorised as wholes and do not participate in inflectional clusters: where are the remaining parts of a partially memorised form to be found in the lexicon? We have no room here to address these questions at the level of detail they deserve. We can only suggest that the maps shown in the present paper represent a (first) level of morphological (as opposed to lexical) organization of the space of inflected word forms. We know that word forms occupy a multidimensional linguistic space, and can thus be classified according to multiple perspectives. In this paper, we were mainly concerned with issues of morphological processing and classification, because of the peculiar and paradoxical problems they seem to raise. No doubt, a full psycho-computational account of the mental lexicon should make provision for several classificatory layers, which, in the present framework, are likely to be associated with separate, independently self-organized, but associatively-related maps.

Another interesting, related issue has to do with the classical dynamics of child morphology learning known, since the seminal work of Rumelhart and McClelland (1986), as the U-shaped curve (Plunkett and Marchman 1991), and its relationship to our computational model. In fact, it would not be too difficult to equate the first phase of SOM learning, where only very frequent items are memorised in full, with the stage of rote memorization characterizing the top left onset of a U-shaped learning curve. The intermediate, over-regularization stage may in turn correspond to a phase where bigger clusters set in, thus pushing itemized learning into the background. Eventually, the final, mature stage may correspond to a phase of learning fine-grained morphological classes. For this picture to be put to the challenging test of a computer simulation, however, several further steps remain to be taken in the direction tentatively suggested here.

References

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

Anderson, J.R., 1993, *Rules of the Mind*. Hillsdale, NJ: Lawrence Erlbaum.

Anderson S., 1992, *A-Morphous Morphology*. Cambridge, UK: Cambridge University Press.

Aronoff M., 1994, *Morphology by itself*. Cambridge (MA): The MIT Press.

Boyland, J.T., 1996, *Morphosyntactic Change in Progress; a Psycholinguistic Approach.* Dissertation: Linguistic Department, University of California.

Bybee, J., 2002, Sequentiality as the basis of constituent structure. in T. Givón and B. Malle (eds.) *The Evolution of Language out of Pre-Language*, Amsterdam: John Benjamins. 107-132.

Bybee J., 1985, *Morphology: A study of the relation between meaning and form*, Benjamins.

Bybee J., 1995, Regular Morphology and the Lexicon, *Language and Cognitive Processes 10* (5).

Caramazza A., Laudanna A., Burani C., 1988, Lexical access and inflectional morphology. *Cognition 28*.

Clahsen, H., 1986, Verb inflections in German child language: acquisition of agreement markings and the functions they encode, *Linguistics 24*. 79-121.

Clahsen H., 1996, Introduction. In H. Clahsen (ed.) *Generative perspectives on language acquisition: Empirical findings, theoretical considerations, and cross-linguistic comparisons*. Benjamins.

Corbett, G., Hippisley, A., Brown, D., 2001, Frequency, regularity and the pardigm: A perspective from Russian on a complex relation, in J. Bybee, P. Hopper (eds.) Frequency and the emergence of linguistic structure, Benjamins.

Croft, W., 2001, Radical Construction Grammar, Oxford University Press.

Creutz, M. and Lagus, K. 2004. Induction of a Simple Morphology for Highly-Inflecting Languages. In Proceedings of the 7th Meeting of the ACL SIGPHON, Barcelona.

Dick, F., Bates, E., Wulfeck, B., Utman, J., Dronkers, N. & Gernbascher, M.A., 2001, Language deficits, localization and grammar: Evidence for a distributive model of language breakdown in aphasics and normals. *Psychological Review, 108(4), 759-788.*

Ellis, N.C., Schmidt, R., 1998, Rules or Associations in the Acquisition of Morphology? The Frequency by Regularity Interaction in Human and PDP Learning of Morphosyntax. *Language and Cognitive Processes 13 2/3*. 307-336.

Frost, R., Forster K.I., Deutsch A., 1997, What can we learn from the morphology of Hebrew? A masked-priming experiment of morphological representation, Journal of Experimental Psychology: Learning, Memory and Cognition 23(4).

Frost, R., Grainger J., Rastle K., 2005, Current issues in morphological processing, Language and cognitive processes 20(1/2).

Gaussier, E., 1999, Unsupervised Learning of Derivational Morphology from Inflectional Lexicons, in Proceedings of Unsupervised Learning in Natural Language Processing, ACL 1999 Workshop, Maryland USA.

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

Goldsmith, J., 2006, An algorithm for the unsupervised learning of morphology. *Natural Language Engineering 12*. 1-19.

Goldsmith, J., O'Brien, J., 2006, Learning inflectional classes. MS.

Goldberg, A.E., Casenhiser, D., Sethuraman, N., 2004, Learning Argument Structure Generalizations, *Cognitive Linguistics*.

Hare M. L., Ford M., Marslen-Wilson W.D., 2001, Ambiguity and frequency effects in regular verb inflection, in J. Bybee, P. Hopper (eds.) Frequency and the emergence of linguistic structure, Benjamins.

Harris, Z., 1951, *Methods in Structural Linguistics*. Chicago: University of Chicago Press.

Haspelmath, M., 2002, *Understanding Morphology*. London: Understanding Language Series, Arnold.

Hay, J., R.H. Baayen, 2005, Shifting paradigms: gradient structure in morphology, *Trends in Cognitive Sciences*, 9(7): 342-348.

Hyams, N., 1986, *Language Acquisition and the Theory of Parameters,* Dordrecht, Reidel.

Hyams, N., 1996, The underspecification of functional categories in early grammars, in H. Clahsen (ed.), *Generative Perspectives on Language Acquisition*, Amsterdam, Benjamin.

Jenkins, W., Merzenich M., Ochs M., 1984, Behaviorally Controlled Diferential Use of Restricted Hand Surfaces Induces Changes in the Cortical Representation of the Hand in Area 3b of Adult Owl Monkeys. *Society for Neuroscience Abstracts 10*. 665.

Juola, P., Plunkett, K., 1999, A Connectionist Model of English Past Tense and Plural Morphology. *Cognitive Science 43(1)*.

Kaas J., Merzenich M., Killackey H., 1983, The Reorganization of Somatosensory Cortex Following Peripheral Nerve Damage in Adult and Developing Mammals. *Annual Review of Neuroscience 6*. 325-56.

Klein, W., C. Perdue, 1997, The Basic Variety (or: Couldn't natural languages be much simpler?), *Second Language Research 13 (4):301-347*.

Kohonen, T., 2001, *Self-Organizing Maps.* Springer.

Leech, G., 1992, 100 million words of English: the British National Corpus. *Language Research 28 1*. 1-13.

Lightfoot, D., 1999, *TheDevelopment of Language. Acquisition, Change, and Evolution*, Blackwell publishers.

Matthews, P.H., 1992, *Morphology*. (second edition) Cambridge UK: Cambridge University Press.

Matthews, P.H., 1993, *Grammatical Theory in the United States*. Cambridge UK: Cambridge University Press.

Marslen-Wilson, W.D., Tyler, L.K., Waksler, R., Older, L., 1994, Morphology and meaning in the English mental lexicon, *Psychological Review 101*. 3-33.

Maamouri, M., Bies, A., Buckwalter, T., Mekki W., 2004, The Penn Arabic Treebank: Building a Large-Scale Annotated Arabic Corpus, NEMLAR International Conference on Arabic Language Resources and Tools: Cairo.

Marcus, G.F., Brinkmann, U., Clahsen, H., Wiese, R. and Pinker, S., 1995, German Inflection: The exception that proves the rule. *Cognitive Psychology 29*. 198-256.

Montemagni, S. et al., 2003, Building the Italian Syn-tactic-Semantic Treebank, in A. Abeillé (ed.), *Building and using Parsed Corpora*. Language and Speech series, Dordrecht: Kluwer. 189-210.

Mowrey R., Pagliuca W., 1995, The reductive character of articulatory evolution, *Rivista di Linguistica 7* (1).

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., 2003, Is machine language learning yet another machine learning task? In Proceedings of *AI*IA*: Pisa.

Pirrelli, V., 2000, *Paradigmi in Morfologia: un approccio interdisciplinare alla flessione verbale dell'italiano*. Pisa-Roma: Istituti Editoriali e Poligrafici Internazionali.

Pirrelli V., Calderone B., Herreros I., Virgilio M., 2004, Non-locality all the way through: Emergent Global Constraints in the Italian Morphological Lexicon , In Proceedings of 7th Meeting of the ACL SIGPHON, Barcelona.

Plaut, D.C., Gonnerman, L.M., 2000, Are non-semantic morphological effects incompatible with a distributed connectionist approach to lexical processing?, Language and Cognitive Processes 15(4/5).

Plunkett, K., Marchman, V., 1991, U-shaped learning and frequency effects in a multi-layered perceptron. Implications for child language acquisition. *Cognition 38*: 3-102.

Plunkett, K., Juola, P., 1999, A Connectionist Model of English Past Tense and Plural Morphology. *Cognitive Science* 43(1).

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

Rissanen, J., 1989, *Stochastic Complexity in Statistical Inquiry*. World Scientific Series in Computer Science. Singapore: World Scientific.

Rumelhart, D., McClelland, J., 1986, On learning the past tense of English verbs. In D.E. Rumelhart and McClelland J. (eds.) *Parallel distributed processing: Explanations in the microstructure of cognition*. Cambridge (MA): MIT Press.

Say, T., Clahsen, H., 2001, Words, Rules and Stems in the Italian Mental Lexicon. In S. Nooteboom, F. Weerman and F. Wijnen (eds.) *Storage and computation in the language faculty*. Dordrecht: Kluwer Academic Publishers. 75-108.

Schone, P., Jurafsky, D., 2000, Knowledge-free induction of morphology using Latent Semantic analysis, in Proceedings of CoNLL 2000 and LLL-2000, Lisbon.

Slobin, D. I., 1997, The Origins of Grammaticalicizable Notions: Beyond the Individual Mind, in Dan I. Slobin (ed.) The Crosslinguistic study of Language Acquisition, Lawrence Erlbaum Associates.

Stemberger J.P. MacWhinney B., 1988, "Are inflected forms stored in the lexicon?" In M. Hammond and M. Noonan (eds.) *Theoretical Morphology. Approaches in modern linguistics*. Academic Press.

Stump G., 2001, *Inflectional Morphology: A Theory of Paradigm Structure*. Cambridge (UK): Cambridge University Press.

Strack F., Mussweiler T., 1997, "Explaining the enigmatic anchoring effect: Mechanisms of selective accessibility" *Journal of Personality and Social Psychology* 73.

Tomasello, M., 2000, Do young children have adult syntactic competence? *Cognition 74*. 209-253.

Tomasello, M., in press, Acquiring linguistic constructions. To appear in R. Siegler, D. Kuhn (eds.), *Handbook of Child Psychology*.

Waibel A., Hanazawa T., Hinton G., Shikano K., Lang K., 1989, "Phoneme Recognition Using Time-Delay Neural Networks" *IEEE Transactions on Acoustics, Speech, and Signal Processing, Vol. 37, No. 3*

Wicentowski R., 2004, Multilingual Noise-Robust Supervised Morphological Analysis using the WordFrame Model, In Proceedings of 7th Meeting of the ACL SIGPHON, Barcelona.

Wilson, S., 2003, Lexically specific constructions in the acquisition of inflection in English. *Journal of Child Language 30*. 75-115.