

MORPHOLOGICAL STRUCTURE THROUGH LEXICAL PARSABILITY

CLAUDIA MARZI

MARCELLO FERRO

VITO PIRRELLI

ABSTRACT: The emergence of morphological structure in lexical acquisition is analysed in the computational framework of Temporal Self-Organising Maps (TSOMs), to provide an explanatory basis for both psycholinguistic and linguistic accounts of lexical parsability. The investigation we propose is grounded on the hypothesis that perception of morphological structure (parsability) and frequency strongly correlate in the acquisition of inflectional paradigms. Analysis of experimental results of word acquisition obtained by artificially varying training conditions, allows us to understand developmental competition between fully-inflected word forms, and to investigate a hierarchy of frequency effects. The computational and theoretical implications of such a memory-based view of the relationship between frequency and perception, and its potential to account for long-term morphological effects in lexical acquisition are illustrated.

KEYWORDS: inflectional paradigms, morphological structure, token/type frequency, word processing.

1. INTRODUCTION

One of the classical assumptions in the psycholinguistic literature on the mental lexicon is that parsed affixes are associated with independently activated access units that tend to spread activation to affix-sharing words, and that activation levels strongly correlate with the affix productivity (Marslen-Wilson et al. 1996; Duñabeitia et al. 2008; Smolík 2010). A number of influential papers (Hay 2001; Hay & Baayen 2002; Hay & Plag 2004) suggested that parsability criteria strictly interact with FREQUENCY to define morphological productivity and word structure constraints in the lexicon. For example, the frequency of a derivative (e.g. *government*) relative to its base (*govern*) is shown to be a good predictor for parsability and productivity. The higher the base/derivative frequency ratio is, i.e. the higher the frequency of a base relative to the frequency of its derivative, the more likely the morphological structure of the latter to be perceived, and the associated affix to be used productively.

In multiple access models of the morphological lexicon (Burani & Laudanna 1992; Baayen et al. 1997), upon presentation of a morphologically complex word form, both the whole-word and sublexical access units compete for activation, as a function of (i) contextual information of recent-

ly activated access units (as in *priming* experiments), and (ii) how often access units have been activated (*token* frequency effects). In the absence of strong contextual effects, if the frequency of the input word is higher than that of its constituent parts, the former is more likely to be accessed as a whole. Otherwise, a decompositional access route is preferred. Accordingly, the high frequency of an affix determines high rates of decompositional access for words containing it, and ultimately ensures the affix productivity. Conversely, it is supposed that an affix contained only by words that access the whole-word route is unlikely to be productive. Once more, these models posit a strong connection between productivity and decompositional parsing in perception and processing.

The correlation between frequency of input forms and perception (or lack of perception) of their structure, shows that it is not possible to decouple representations from processing operations. Access representations in the lexicon may differ exactly because words are differentially processed in serial perception and storage. For a word like *government* to be mapped onto two access units (*govern* and *-ment*), these units must be perceived and stored independently. This does not only imply a parsing stage where the input word is split into its parts and mapped onto the corresponding access units. It also presupposes a perceptual alignment between the lexical representation of *government* and the representations of other derivatives sharing the same affix, for them to be perceived and recoded in terms of partially overlapping access units.

The interplay between frequency and perception has been the focus of intense investigation in the literature on working memory and the human ability to recode and retain sequences of linguistic items (e.g. letters, segments, syllables, morphemes or words) (Gathercole & Baddeley 1989; Papagno et al. 1991; Baddeley 2007). Items that are frequently sequenced together are known to be stored in long-term memory as single chunks, and accessed/executed as having no internal structure. This increases fluency, eases comprehension and also explains the possibility to retain longer sequences in short-term memory when familiar chunks are presented (see Cowan 2001, for a detailed overview). Memory processes for serial cognition are helpful in establishing the explanatory link between the developmental course of word memory traces in the mental lexicon and their organisation and role in word perception, access and productivity. More interestingly for our present concerns, parts belonging to high-frequency chunks tend to resist being perceived as autonomous elements and thus being used independently.

In this paper, we will investigate the computational implications of such a memory-based view of the relationship between frequency and perception, and its potential to account for long-term morphological effects in lex-

ical ACQUISITION. In particular, we will illustrate how high frequency yields the WHOLENESS (lexical entrenchment) effect, why frequently-used words compete with members of their own lexical families (inflectional paradigms in our present work) and why they hardly participate in larger series of morphologically homologous words (e.g. any paradigmatic cell of an inflectional family).

In the computational framework we propose here (Temporal Self-Organising Maps or TSOMs for short), word processing and lexical acquisition are implemented as recoding and storage strategies of time-series of symbols, dependent on language-specific factors and extra-linguistic cognitive functions such as lexical organisation, lexical access and recall, input-output representations, and memory self-organisation. TSOMs provide a general framework for putting algorithmic hypotheses of the processing-storage interaction to the empirical test of a computer implementation. In the ensuing sections, we first outline the general architecture of a TSOM, to then focus on a few implications of this view from the perspective of emergent lexical representations and word processing.

2. TSOMs

In TSOMs (Figure 1, left), a variant of Kohonen's Self-Organising Maps (Kohonen 2001), classical input connections (or WHAT connections) conveying the current input stimulus to each map node are augmented with re-entrant Hebbian connections (or WHEN connections) to encode probabilistic expectations over incoming serial stimuli (Koutnik 2007; Ferro et al. 2010; Pirrelli et al. 2011; Ferro et al. 2011; Marzi et al. 2012a, 2012b). Connections are functional equivalents of neuron synapses, and store weights in the 0-1 interval, determining the amount of influence that the firing (ACTIVATION) of one node has on another node. Map nodes are fully inter-connected by two bundles of Hebbian connections: in-going (or pre-synaptic) connections and out-going (or post-synaptic) connections.

At their core, TSOMs are dynamic memories that are trained to store and classify input stimuli through PATTERNS OF ACTIVATION of map nodes, also referred to as MEMORY TRACES. A pattern of node activation is the processing response of a map when a stimulus is input. After being trained on a set of stimuli, the map learns to respond to similar stimuli with largely overlapping activation patterns.¹ In the end, a memory trace of a stimulus on the map consists in the dynamic, successful processing response of the map when the stimulus is shown.

¹ The distance between any two patterns is measured as the mean per-node difference between their levels of activation (or CO-ACTIVATION DISTANCE).

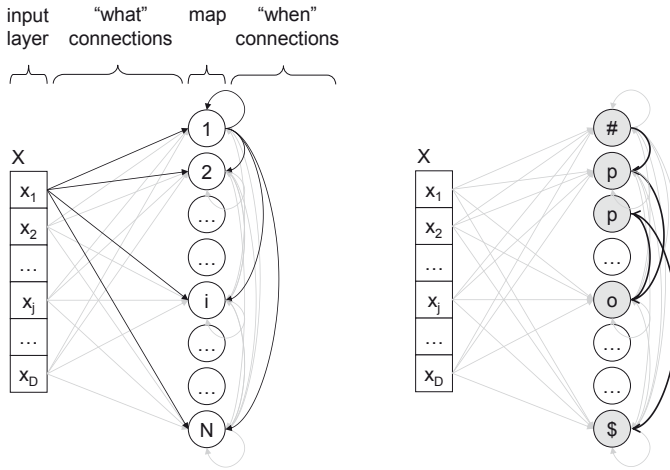


FIGURE 1. (LEFT) OUTLINE ARCHITECTURE OF A TSOM.
 (RIGHT) BMU ACTIVATION CHAIN FOR THE INPUT STRING “#POP\$”.

This behaviour unfolds through a short-term and a long-term dynamic. In the short-term, at time t an input stimulus is shown to the map by being encoded on the input layer (Figure 1, left). Information propagates through WHAT connections to map nodes, which are activated synchronously as a function of the STRENGTH of connection weights. Meanwhile, the same nodes receive re-entrant signals through WHEN connections, propagating the level of activation of nodes responding to the immediately preceding stimulus (time $t-1$). Competition among co-activated nodes is won by the most highly activated node, or Best Matching Unit (BMU) at time t .

Such a short-term activation is followed by a long-term training dynamic. Both WHAT and WHEN connections of the current BMU are potentiated, for the latter to be more responsive when the same input stimulus is presented to the map over again. In addition, weight potentiation spreads to neighbouring nodes of the current BMU, so that topologically close nodes on a map get increasingly more sensitive to stimuli that are found similar on WHAT and WHEN levels of connectivity. At the same time, nodes that are not topologically close to the current BMU are inhibited.

Such a dynamic specialisation has interesting consequences on the way words are represented on the map. For a map, words are sequences of symbols, beginning with “#” and ending with “\$”, that are shown one after the other on the input layer in their left-to-right order. At each time tick, the current word symbol activates one BMU and the process is repeated for each symbol until the end-of-word symbol (“\$”) is reached. The map’s activation state is eventually reset upon presentation of a new word (signalled by the appearance of “#” on the input layer).

An input word thus leaves on a map a temporal chain of consecutively activated BMUs, representing the way the word is internally recoded by the map. This is shown in Figure 2 (right), depicting an ACTIVATION CHAIN of BMUs after the string “#pop\$” is input. It should be noted that two distinct (but topologically neighbouring) nodes respond to the two *p*'s in *pop*, bearing witness to the process of selective sensitivity to time-bound instances of the same symbol type. We consider the implications of this dynamic on lexical representations and word processing in the ensuing section.

3. LEXICAL REPRESENTATIONS AND PROCESSING

Emergentist, associative views on the morphological lexicon (Bybee 1995, 2002; Burzio 2004; among others) treat word forms as primitive units and their recurrent parts as derivative abstractions over word forms. According to this perspective, full forms constitute the basis for morphological processing, with sub-lexical units resulting from the application of morphological processes to full forms. Morphology acquisition relies on the emergence of the relations between fully stored word forms, which are concurrently available in the speaker's mental lexicon and jointly facilitate processing of morphologically-related forms.

In a network-based interpretation of the associative view (Bybee 1995) word forms sharing meaning components and/or phonological/orthographic structure are associatively connected with one another, as a function of FORMAL TRANSPARENCY, TOKEN FREQUENCY, and size of morphological family (TYPE FREQUENCY). In particular, the strength of lexical connections is affected by frequency. High-frequency word forms have greater lexical autonomy: their lexical connections with other morphologically related forms are weaker. Hence, the strength of a pattern is inversely proportional to the number of times a particular sequence of symbols (a full form, a stem or an affix) instantiates the pattern. On the other hand, the strength is directly proportional to the number of different word types where the sequence is found.

Bybee's network model resorts to two different mechanisms, lexical entrenchment and lexical association, to account for the inverse correlation between frequency and lexical productivity. Other symbolic data structures, such as WORD TREES and WORD GRAPHS, make use of lexical connections between consecutively occurring symbols to model both entrenchment of individually stored forms and associative relations between concurrently stored forms.

In a tree-like representation (Figure 2, top), redundant patterns are represented by shared chains of connected nodes, where nodes stand for basic representational symbols (e.g. letters or sounds), and directed (forward) arcs link two consecutively occurring symbols.

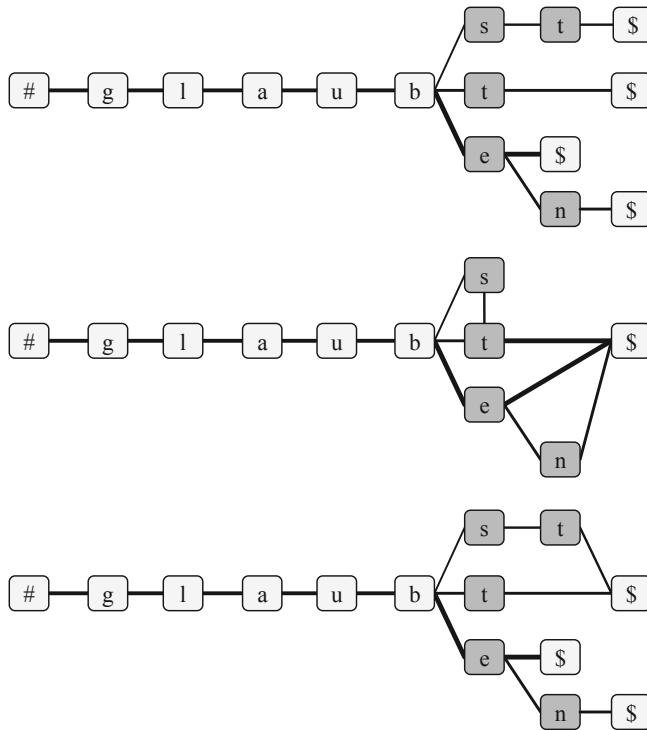


FIGURE 2. TREE-LIKE (TOP) AND GRAPH-LIKE (MIDDLE AND BOTTOM) REPRESENTATIONS OF FOUR GERMAN VERB FORMS.

In a probabilistic interpretation of a word tree, the strength of each connection reflects how often the symbols corresponding to connected nodes are seen one after the other in the input data. Hence, a high-frequency form tends to develop a chain of strongly connected nodes. The strength of connections defines the level of entrenchment of that form in the lexicon and can be interpreted dynamically as the conditional probability with which a particular form is expected to occur, when an increasingly longer part of the word is perceived. In the tree, word forms belonging to the same paradigm (e.g. *glaubst* / *glaube* / *glauben* ‘you / I / we-they believe’, and *glaubt* ‘he-she believes’) share a chain of nodes representing the common stem. The entropy of the forward conditional probability distribution of the paths branching out of the same stem is a measure of the level of indecision between possible alternative continuations of the same stem, and ultimately reflects the amount of structure shared by the forms belonging to the same paradigm.

A more compact representation of the same set of word forms is provided by word graphs (Figure 2, middle). Whenever possible, word forms

of the same paradigm share part of the suffix nodes (e.g. *glaube* ‘I believe’, and *glauben* ‘we/they believe’, or *glaubst* ‘you believe’, and *glaubt* ‘he-she believes’). The structure allows any node to be reached by more forward connections at the same time. This produces a considerable reduction in the number of nodes needed to represent morphologically-related forms. The entropy of the backward conditional probability distribution of all paths branching into the same ending is thus proportional to the number of word forms that share the ending, and ultimately reflects the level of common structure shared by sets of paradigms (conjugation classes).

Word trees and word graphs need not be considered as mutually exclusive data structures. Figure 2 (bottom) shows a somewhat intermediate data structure, representing activation chains on a map, with squares representing BMUs, and arcs representing connections between consecutively activated BMUs. In previous work (Marzi et al. 2012a, 2012b, 2012c), we showed that the extent to which TSOMs organise morphologically-related words through either word trees or word graphs depends on the degree of probabilistic support received by the network of long-term associative relations among all stored word forms. High-frequency forms are more likely to be represented as word trees, with little or no sharing of recurrent endings with other word forms. Conversely, low- and medium-frequency forms have a tendency to share endings. Evenly-distributed word frequencies are in fact translated into evenly-distributed weights over temporal connections, i.e. into unbiased (more uniform) expectations over a wider range of possible graph continuations. By analogy to forward and backward probability distributions, the entropy of the distribution of (normalised) weights over post-synaptic (i.e. branching-out) and pre-synaptic (i.e. branching-in) connections gives a measure of the level of co-activation among forms belonging to the same paradigm (e.g. *gibt* and *gibst*), and among forms selecting the same ending in different paradigms (e.g. *machen* and *finden*) respectively.

TSOMs can thus simulate memory-based, self-organisation processes leading to acquisition of a graded composition of tree-like and graph-like data structures, as a function of degrees of memory entrenchment. They are in line with theoretical models of emergent lexical organisation (e.g. Burzio 2004) and neuro-functional architectures of the language processor (e.g. Catani et al. 2005) that blur the distinction between storage and computation, along with the dichotomy between morphological representations and morphological rules (Bates & Goodman 1997, 1999). In fact, morphological representations are determined by probabilistically recurrent processing strategies cached in long-term memory.

In TSOMs, concurrently stored forms compete for the same pool of memory nodes due to the interplay of several factors: (i) number of available memory nodes (relative to the number of input words); (ii) WORDLIKE-

NESS of input words; (iii) token frequency of input words; (iv) plasticity of the map. When a wordlike input, i.e. a sequence of symbols that is found in (many) other words in the lexicon (Bailey & Hahn 2001), is shown to a map, it will fail to develop an exclusively dedicated chain of nodes, due to the amount of shared formal redundancy. At the same time, if an input word is not wordlike, it is more likely to activate a dedicated node sequence. Finally, plasticity defines the map's readiness to adjust connection weights. During training, the map loses its plasticity, so weights are adjusted less and less adaptively as training progresses (for a detailed description of temporal layer plasticity and long-term potentiation mechanisms see Pirrelli et al. 2011; Marzi et al. 2012a).

In this perspective, CO-ACTIVATION of the same BMUs by different input words reflects the extent to which the map perceives surface morphological relations between fully-stored words, and represents, at the same time, a logical precondition to morphological generalisation. PERCEPTION of surface morphological relations presupposes that the map can coactivate already stored, recoded representations of morphologically-related input words. Morphological GENERALISATION, on the other hand, describes the process whereby a novel form (e.g. a yet unknown inflected form of a known lexical exponent) is recoded on the basis of activation of other forms. The more morphological relations are perceived, the better the KNOWLEDGE of a map, and the more likely the resulting generalisations will be.

4. FREQUENCY AND THE DYNAMIC OF PARADIGM ACQUISITION

The time course of lexical acquisition is known to be affected by several factors, ranging from word length, word frequency and time of acquisition, to wordlikeness and perceptual salience. In particular, token frequency is understood to facilitate lexical access and correlate negatively with response latencies in visual lexical decision (Taft & Forster 1975; Whaley 1978). High-frequency words tend to be acquired earlier and this may later impact on lexical tasks due to loss of plasticity (Ellis & Morrison 1998; Zevin & Seidenberg 2002). High token frequency is traditionally understood to protect irregulars from regularisation effects through time (Bybee 1985, 1995; Corbett et al. 2001), and to leave deeply entrenched memory traces in the mental lexicon (Alegre & Gordon 1999; Baayen et al. 2007). Accordingly, high-frequency word forms can be learned by rote and tend to show greater lexical autonomy, even from other members of their own paradigms (see Bybee 1995), thus weakening their lexical connections with shared morphological schemata. This explains the salience of a whole relative to its parts

in high-frequency morphologically complex words (Hay & Baayen 2002) and lack of productivity and overall resistance to change of affixes embedded in high-frequency words (Bybee 1985, 1995; Hay 2001; Corbett et al. 2001). Finally, recent evidence has shown the role of frequency distributions in morphological families. Uniform frequency distributions over members of the same inflectional paradigm (measured in terms of inflectional entropy) make members more readily accessible (Moscoso et al. 2004a), favouring a better allocation of memory resources.

All this evidence appears to support a view of the mental lexicon as a dynamic integrative system, whereby words are concurrently, redundantly and competitively stored. The main consequences of such a view are clear in broad outline, but some of their implications have been questioned on several grounds, especially in connection with the role of frequency factors in affecting acquisition of regular and irregular paradigms. Furthermore, no existing model of lexical processing can account for all these facts.

For associative, connectionist models of lexical memory, all inflected lexical items should be sensitive to token frequency, irrespective of their degrees of inflectional regularity (Plunkett & Marchman 1993; McClelland & Patterson 2002; Bybee 1985; 1995; 2002). The only difference between more and less regular inflectional classes is accountable in terms of frequency and, ultimately, memory effects (Ellis & Schmidt 1998). Stems are regular if they show up in many paradigm cells, and endings are productive if they are found in a large family of forms instantiating the same paradigm cell. Productivity, in both cases, is governed by type frequency, or, equivalently, by the size of the relevant word family. Finally, acquisition and productivity are both favoured by wordlikeness, i.e. by the extent to which a form is perceived as similar to other forms in the lexicon. Accordingly, endings are generalised more easily if they are used in paradigms with a high family resemblance.

On the contrary, dual mechanism models (Pinker 1999; Pinker & Ullman 2002) assume that regular forms are computed on-line, and should thus be insensitive to both token and type frequency effects. Similarly, regular endings are categorically distinct from other morphological processes. They tend to be mastered earlier than irregular endings, and are typically applied (and possibly over-applied) in an all-or-none fashion, with no correlation with type frequency effects. A logical implication of these two assumptions is that acquisition of a regular inflectional paradigm should be a sudden event: once its stem/root is committed to memory and the relevant endings are mastered, all its members should be acquired instantaneously. This is also true of any regular inflectional ending, which is generalised neither as a function of the number of words where it is found, nor on the basis of their degree of family resemblance. On the other hand, irregular

paradigms are expected to be acquired in a piecemeal, item-based fashion and should thus be sensitive to cumulative token frequency effects in acquisition accuracy. In the end, irregular inflections are rarely generalised. If they can be generalised at all, this is caused by formal resemblance.

For such a complex range of issues to be settled, psycholinguistic evidence is simply not enough. Investigating the dynamic interaction between type and token frequency in overlapping patterns of lexical activation that evolve with time requires careful control of all the quantitative variables involved, with little if any change in the range and nature of the lexical data used for training. This is virtually impossible to obtain in longitudinal studies or through experimental protocols involving human subjects. Computer modelling is useful in this connection for several reasons. First, it allows direct manipulation of frequency and timing of exposure to word stimuli that are directly controlled under possibly artificial training conditions. Second, it implements an explicit model of memory and acquisition that allows differential effects of various training conditions to be observed and assessed quantitatively. Finally, it provides a truly explanatory model of the evidence observed, as this evidence is the replicable outcome of the dynamic interaction of a few basic processing principles.

4.1 The role of input frequency

To address issues of frequency – both token and type frequency – and developmental acquisition, we ran a series of simulations, designed to investigate the interconnection between time of acquisition, frequency distribution and regularity of inflectional paradigms.

We selected from CELEX (Baayen et al. 1995) the top 50 high-frequency German verb paradigms (34 irregular/strong paradigms and 16 regular/weak paradigms). From each paradigm, 15 inflected forms were extracted: the full set of present indicative (6) and präteritum (6) forms, the past participle, the infinitive and the present participle. All 750 forms were encoded as strings of letters, starting with “#” and ending with “\$”. All letters were encoded on the map’s input layer as mutually orthogonal binary vectors. Each input word was administered to a 40 x 40 nodes map one letter at a time, with memory of past letters being reset upon presentation of the start-of-word symbol “#”.

To factor out frequency effects and assess the role of frequency distributions vs. absolute frequencies, we trained identical maps on the same dataset administered under two distinct regimes. In one training regime, input forms were repeatedly presented to the map as a function (into the 1-1001 range) of their frequency distribution in CELEX, for a total number of 10,286 token presentations per learning epoch (skewed training protocol). In the second training regime, the same set of input forms were shown to

the map 5 times each, for a total number of 3,750 presentations per epoch (uniform training protocol). In both regimes, the order of word presentation was randomised and results were averaged over five instances of the same map trained under each training regime for 100 epochs.

We tested the two groups of maps (hereafter referred to as skewed maps and uniform maps) on the task of WORD RECALL, and compared their behaviour through the time course of lexical acquisition. Word recall simulates retrieval of a sequence of letters by letting the map go through a BMU activation chain such as the one for *pop* illustrated in Figure 1 (right), and iteratively output, at each time tick, the symbol associated with the current BMU. A BMU at time t is calculated by overlaying the activation chain with the expectations of the BMU at time $t-1$, propagated through WHEN connections only (i.e. in the absence of any stimulus presented on the input layer). The process is iterated until a “\$” is output.

Errors occur when the map misrecalls one or more symbols in the input string, by either replacing them with different symbols or by outputting correct symbols in the wrong order. Partial recall, i.e. the correct recall of only a substring of the target word (e.g. “#GEBENS\$” for “#GEGEBENS\$”), is counted as an error as well.

The incremental time course of word acquisition for the two groups of maps is plotted in Figure 3 (top). For any word form, we define its time of acquisition as the first epoch starting from which the form is recalled correctly. Unlike word recognition, which mostly depends on the current input stimulus, word recall depends on internal recoding only, and requires that fine-grained information about nature and timing of the letters making up the word is stored in the internal state of the map. The plot shows how many words are acquired per learning epoch, as a percentage of all input words. Counts are averaged over five map instances for each training conditions, with standard deviation represented by whiskers. Clearly, results are shown by both word types and word tokens for skewed maps only.

As a general trend, skewed maps are found to acquire word types more slowly than uniform maps², particularly in the 12-20 epoch range, suggesting that there is a statistically significant advantage ($p < .001$ at epoch 20) in having all forms presented an equal number of times during training. As the training algorithm minimises the probability for a map to make a recognition error, a map trained on skewed distributions will tend to acquire the most frequent forms first, thereby neglecting less frequent ones. This is shown by the mean frequency of correctly recalled forms at each epoch (Figure 3, bottom), and by the statistically significant inverse correla-

² The advantage is significant even when we compare those word forms only that are presented an equal number of times (5) in both training conditions.

tion ($p < .005$) between word token frequency and epoch of word acquisition by skewed maps. Since top-frequency words are not necessarily the most wordlike in the training set (i.e. the most similar to the majority of words), recruitment of memory resources for storing high-frequency items may effectively slow down lexical acquisition overall.

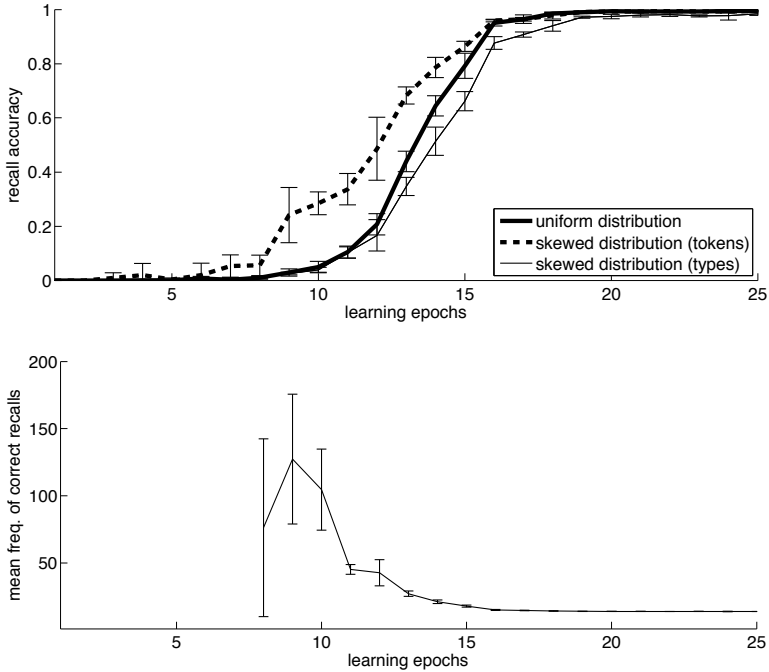


FIGURE 3. (TOP) THE TIME COURSE OF LEXICAL ACQUISITION (RECALL ACCURACY) ON UNIFORM MAPS (THICK SOLID LINE) VS. SKEWED MAPS FOR BOTH TYPE COUNTS (THIN SOLID LINE) AND TOKEN COUNTS (DASHED LINE). (BOTTOM) AVERAGE FREQUENCY OF CORRECTLY RECALLED WORDS BY LEARNING EPOCHS FOR THE SKEWED TRAINING REGIME.

Notably, paradigm regularity is observed to interact with word token frequency. Maps do not memorise words in isolation but in formally-related word families. The impact of word token frequencies on the time course of lexical acquisition thus varies depending on the comparative amount of formal redundancy shared by input words. As shown in Figure 4 (top), the overall growth rate of German word tokens is mostly due to words in irregular (top panel) rather than regular (bottom panel) paradigms. The statistically significant correlation between token frequency and word learning epoch for the whole dataset increases in significance ($p < .001$) for words in irregular paradigms. On the other hand, acquisition of words in regular paradigms is less affected by word token frequency (statistically non-significant correlation). The advantage of having word forms presented

more often to a map thus appears to diminish in regular paradigms, where lexical acquisition has a quicker time course, as confirmed by the steeper rate of acquisition of regular word types, compared with irregular word types, in the 12-20 epoch range.

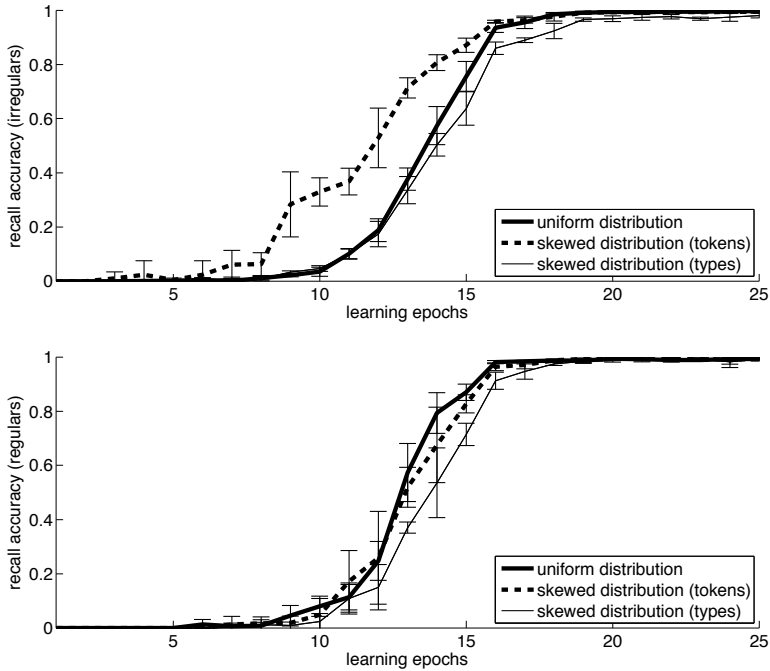


FIGURE 4. THE TIME COURSE OF WORD ACQUISITION (RECALL ACCURACY) OF IRREGULAR (TOP) AND REGULAR PARADIGMS (BOTTOM) IN UNIFORM MAPS (THICK SOLID LINE) VS. SKEWED MAPS FOR BOTH TYPE COUNTS (THIN SOLID LINE) AND TOKEN COUNTS (DASHED LINE).

This can be interpreted once more as a memory effect, since regularity and type frequency are highly correlated in the lexicon. Regular morphological patterns are systematically repeated both intra- and inter-paradigmatically thus being found in many different word types. Irregularly-inflected forms, conversely, are more isolated across paradigm cells and are more dependent on word token frequency. In this context, an interesting related issue is whether there is a direct connection between the time course of word acquisition and the time course of paradigm acquisition.

To answer this question, for each paradigm we define its time of acquisition as the mean acquisition epoch of all forms belonging to the paradigm. The paradigm acquisition epoch provides an estimate of the average time it takes for all forms of the paradigm to be recoded in a time-sensitive way, for them to be recalled accurately from their corresponding activation chains. In Figure 5 (left), we plotted the acquisition epoch of each German

verb paradigm over both training conditions: skewed (black circles) and uniform (white circles). On the vertical axis, paradigms are arranged by increasing acquisition epochs in the skewed training regime, with cumulative paradigm token frequencies shown in brackets. The right-hand regression plot in Figure 5 shows a statistically significant inverse correlation ($p < .0005$) between cumulative paradigm frequencies (on a logarithmic scale) and the epoch of acquisition of paradigms in the skewed training regime. This seems to suggest that paradigms, like words, are acquired by (cumulative) token frequency of their members. In particular, paradigms with low token frequency members are more difficult to acquire, as memory takes time. But the overall picture is considerably more complex than this.

We can understand more of this dynamic if we compare the acquisition rate of the same paradigms in the two training regimes by looking at the spread between paradigm learning epochs in uniform and skewed training regimes (respectively white and black circles in Figure 5, left).

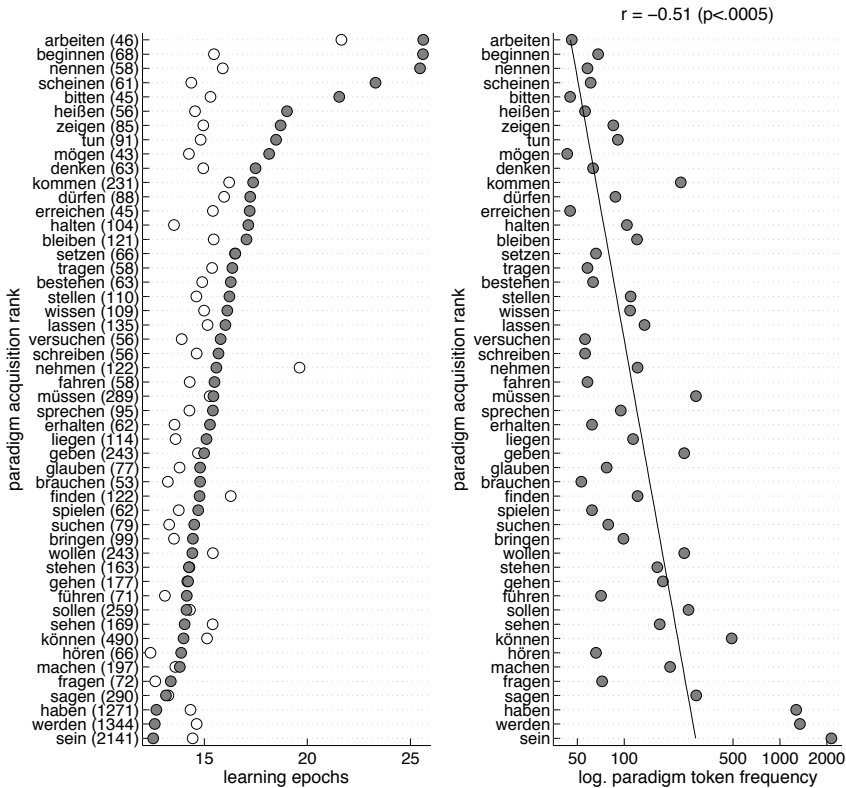


FIGURE 5. (LEFT) ACQUISITION TIME OF PARADIGMS RANKED BY INCREASING LEARNING EPOCH FOR SKEWED (BLACK CIRCLES) AND UNIFORM (WHITE CIRCLES) DISTRIBUTIONS. (RIGHT) REGRESSION OF PARADIGM LEARNING TIME OVER PARADIGM TOKEN FREQUENCY (ON A LOG SCALE).

First, the actual timing of paradigm acquisition appears to depend more on relative frequencies than on absolute frequencies. The vast majority of paradigms are acquired earlier in a uniform training regime ($p < .005$) than when they are presented with skewed frequencies. In addition, word forms in each paradigm are acquired at a quicker rate when they are distributed uniformly: the number of epochs it takes to complete the acquisition of a paradigm after the first member of the paradigm is acquired (or PARADIGM ACQUISITION SPAN) is significantly shorter in uniformly-trained maps ($p < .005$).

A notable exception to this general trend is represented by a few highly irregular paradigms (e.g. SEIN ‘be’, WERDEN ‘become’) that are presented in training with high cumulative token frequencies. This evidence points to a nontrivial interaction between frequency and (ir)regularity. In highly-irregular paradigms with extensive unpredictable stem alternation, relatively isolated stems are acquired in a piecemeal fashion. Unlike widely-distributed, more predictable stems, alternating stems are found in fewer cells of the paradigm (and in some cases in one cell only) and can take little (or no advantage) of cumulative frequency effects across cells. High token frequency can tip the balance in favour of early memorisation, allowing irregular stems to successfully offset their low type frequency. This is what happens with SEIN ‘be’ and WERDEN ‘become’. In medium to low frequency irregular paradigms, however, this is more difficult to happen, and piecemeal memorisation takes longer to set in. This memory effect also explains why, in the skewed training regime, the significant (inverse) correlation between paradigm token frequency and learning epoch is limited to irregular paradigms, while losing significance with regular paradigms (Figure 6). Token frequency effects are more hardly detectable when it comes to acquisition of regular paradigms, simply because predictable stems systematically appear in different forms of the paradigm and can take advantage of their cumulative frequency.

It remains to be understood, however, why uniform distributions determine a clear advantage in time of acquisition, and why the advantage is unevenly apportioned between regular and irregular paradigms, as shown in Figure 7. To shed light on this global interaction, we have to look at the way (ir)regularity and frequency affect the developmental dynamic of memory organisation. The variable patterns of connectivity between recruited nodes, their levels of competition and co-activation on the map can give explanatory insights into the time course of paradigm acquisition and into degrees of perception of structure in the morphological lexicon. From this fine-grained perspective, morphological structure can be investigated as the emergent property of a densely interconnected pool of nodes, whose global behaviour is governed by local patterns of connectivity.

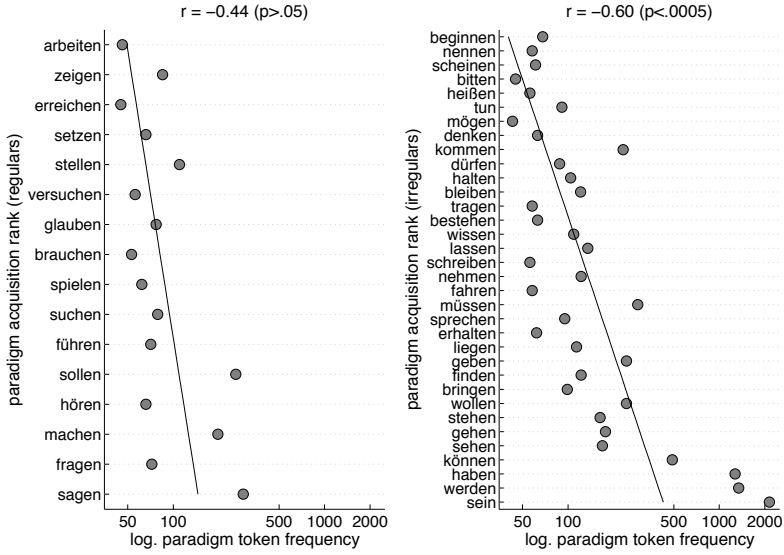


FIGURE 6. REGRESSION OF PARADIGM LEARNING TIME OVER PARADIGM TOKEN FREQUENCY (ON A LOG SCALE) FOR REGULAR (LEFT) AND IRREGULAR (RIGHT) PARADIGMS.

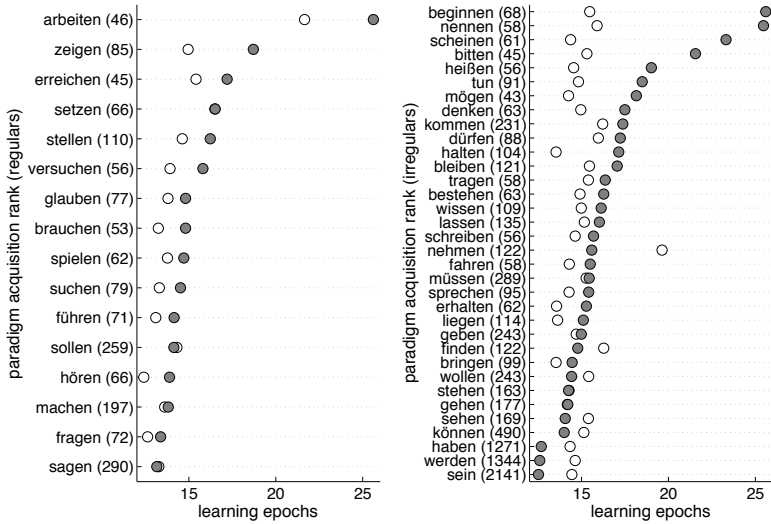


FIGURE 7. ACQUISITION TIME OF REGULAR (LEFT) AND IRREGULAR (RIGHT) PARADIGMS RANKED BY INCREASING LEARNING EPOCH FOR SKEWED (BLACK CIRCLES) AND UNIFORM DISTRIBUTIONS (WHITE CIRCLES).

4.2 *The dynamic of paradigm acquisition*

We already showed (Figure 3, bottom) that maps trained on realistically-distributed data tend to acquire the most frequent word forms first. This is particularly evident in early learning epochs, when individual forms are learned by rote, and the relations between stored word forms are still to emerge. Such an early advantage in WORD ACQUISITION is carried over unevenly to PARADIGM ACQUISITION as training progresses (Figure 5 right, and Figure 6): the strong inverse correlation between paradigm frequency and time of acquisition holds for irregulars ($r=-.60$, $p<.0005$, see Figure 6, right), but gets weakened and statistically not significant with regular paradigms. In addition, this evidence should be contrasted with the rate of paradigm acquisition by maps trained on uniform distributions (white circles in Figure 5, left). When words are administered to a map an equal number of times, paradigms are learned consistently earlier (earlier learning epoch) and more quickly (shorter learning span) than when words are input following skewed distributions. Finally, although regular paradigms are, on average, acquired more quickly than irregular paradigms, they are also less sensitive to differences in frequency distributions: the advantage of having paradigms shown with a uniform distribution (as opposed to a skewed distribution) is considerably smaller when a paradigm is regular.

We suggest that this behaviour is due to the interplay of three factors. First, as observed in the previous section, regular paradigms are formally highly redundant families, where an invariant stem is shared by all paradigm members. Secondly, in our training set regular paradigms tend to be more wordlike than irregular paradigms, meaning that they exhibit significantly more typical (i.e. shared by many word types) word-internal chunk types (trigrams) than irregular paradigms do ($p<.0005$). It is easier for word forms in a regular paradigm to find many other similar forms in the training set, and this is an important prerequisite for knowledge of an inflectional paradigm to be transferred (generalised) to another paradigm.³ Finally, uniformly-trained maps are able to better organise stored words in a deeply interconnected network of associative relations, where more nodes share information through distributed patterns of comparatively poorly entrenched connections. Conversely, high-frequency entrenchment favours individual access and holistic perception, and disfavors co-activation (i.e. spreading of activation to other neighbouring/similar forms) and perception of internal structure.

³ This is confirmed by a significant inverse correlation ($r=-0.64$, $p<.0001$) between paradigm wordlikeness and paradigm acquisition span. That regular paradigms distribute more diffusely in the similarity space of the German verb system also accounts for their apparent insensitivity to strong family resemblance effects.

Due to the interplay of these factors, we conjecture that the across-the-board acquisitional advantage of uniformly-trained maps is a processing effect. Maps are quicker learners of uniformly-distributed data simply because they get more sensitive to patterns of morphological redundancy and are more prone to transfer this knowledge across paradigms. In other words, they are better at generalising knowledge. The advantage between uniform distributions and skewed distributions is less evident in regular paradigms since they are easier to generalise than irregular paradigms in the first place.

To verify sensitivity to morphological redundancy, i.e. evidence of perception of morphological structure by our maps, it is useful to turn back to the lexical representations of Figure 2. The defining feature of tree-like structures (Figure 2, top) is that each memorised word form selects its own distinct ending, with no single node being reached by more than one in-going arc. This amounts to a holistic representation of a word in the lexicon, where the backward probability of finding an in-going arc given a node is always equal to unity. High-frequency words tend to be represented holistically by a map.

In word graphs, the same probability is a function of the number of word paths sharing that node. To estimate the perception of shared structure by a map, we can measure, for each node, the entropy of the distribution of (normalised) weights over its in-going (pre-synaptic) connections, i.e. how likely it is for that node to be reached by any activation chain where it belongs. If a node is strongly selected by one particular chain and weakly selected by other chains, pre-synaptic connection entropy goes down. On the other hand, uniformly-distributed, formally-related words showing a regular morphological pattern tend to share overlapping node chains, and to be densely interconnected with other chains through highly-entropic bundles of pre-synaptic connections. The plot in Figure 8 shows a significantly higher level of pre-synaptic entropy ($p < .0005$ starting from learning epoch 12) for any node in uniform maps when compared with skewed maps.

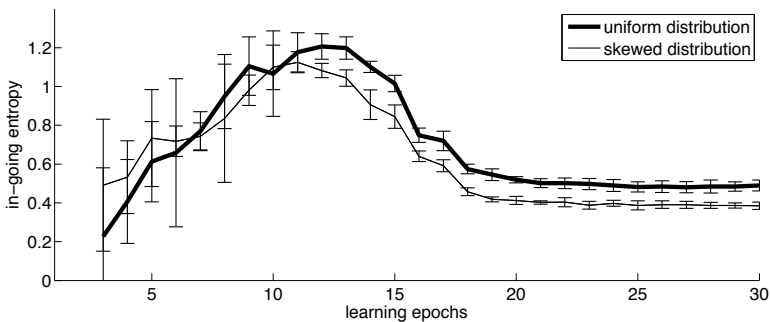


FIGURE 8. AVERAGED ENTROPY OF WEIGHTS ON PRE-SYNAPTIC CONNECTIONS ON BOTH SKEWED (THIN LINE) AND UNIFORM (THICK LINE) MAPS.

We also evaluated the distribution of weights on pre-synaptic connections of nodes that initiate an inflectional ending, i.e. connections crossing a morpheme boundary. We observe no significant difference for the two training regimes when considering morpheme boundaries in low frequency word forms only (i.e. words administered with a token frequency < 5 in the skewed training regime). Statistical significance progressively rises for word sets of increasing token frequencies (Table 1), as shown in Figure 9 for high-frequency forms (token frequency > 50).

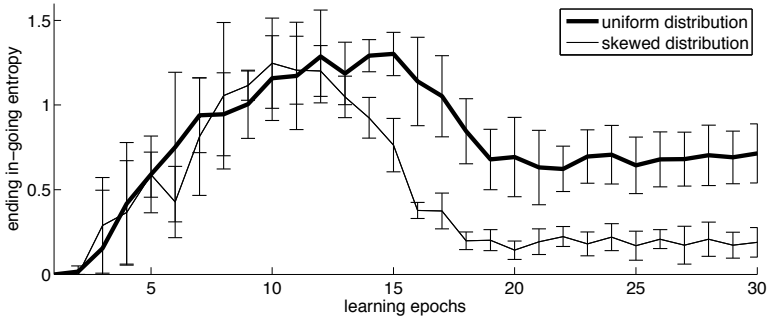


FIGURE 9. AVERAGED ENTROPY OF WEIGHTS ON PRE-SYNAPTIC ENDING CONNECTIONS FOR HIGHLY-FREQUENT FORMS ON SKEWED MAPS (THIN LINE) AND FOR THE SAME WORDS ON UNIFORM MAPS (THICK LINE).

Finally, the strength of connection weights between node chains shows a reversed pattern in the two training regimes (Figure 10). This is intuitively obvious, as connection weights represent a map's expectation over incoming symbols: more deeply entrenched words should thus correlate with stronger expectations.

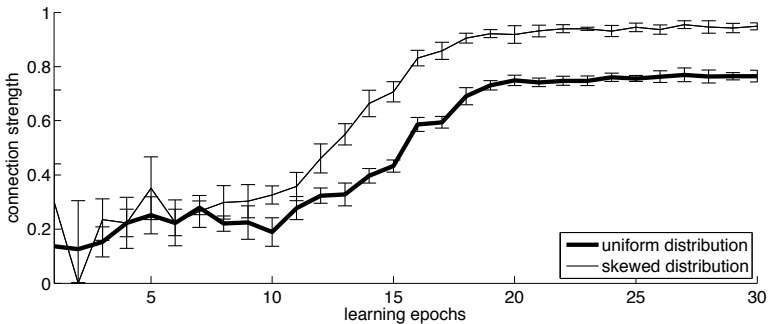


FIGURE 10. AVERAGED STRENGTH OF CONNECTIONS FOR HIGHLY-FREQUENT FORMS ON SKEWED MAPS (THIN LINE) AND FOR THE SAME WORDS ON UNIFORM MAPS (THICK LINE).

Once more, statistical significance progressively rises for word sets of increasing token frequencies (see Table 1). In fact, map connections are known to inhibit each other inter-paradigmatically due to entrenchment (Marzi et al. 2012b): if a suffix node chain is frequently activated by a given stem, it will be far less activated by another stem (in the limit, the strength of all other connections goes down to nil).

| F_{TH} | ENDING IN-GOING ENTROPY | | CONNECTION STRENGTH | |
|----------|---------------------------|----------------|---------------------------|----------------|
| | $F < F_{TH}$ | $F > F_{TH}$ | $F < F_{TH}$ | $F > F_{TH}$ |
| 5 | no statistical difference | U>S (p<.005) | U>S (p<.001) | U<S (p<.0005) |
| 10 | no statistical difference | U>S (p<.005) | U>S (p<.001) | U<S (p<.00005) |
| 50 | U>S (p<.05) | U>S (p<.0005) | U>S (p<.05) | U<S (p<.00005) |
| 100 | U>S (p<.05) | U>S (p<.00005) | no statistical difference | U<S (p<.00005) |

TABLE 1. STATISTICAL SIGNIFICANCE OF THE DIFFERENCE IN ENDING IN-GOING ENTROPY AND CONNECTION STRENGTH BETWEEN UNIFORM (U) AND SKEWED (S) MAPS FOR VARYING FREQUENCY THRESHOLDS. LEVELS OF ENTROPY AND CONNECTION STRENGTH ON SKEWED MAPS ARE CALCULATED FOR WORDS WHOSE FREQUENCY (F) FALLS BELOW ($F < F_{TH}$) OR ABOVE ($F > F_{TH}$) A GIVEN FREQUENCY THRESHOLD F_{TH} , AND COMPARED WITH THE CORRESPONDING VALUES OF THE SAME WORDS ON UNIFORM MAPS. U>S MEANS THAT VALUES ARE SIGNIFICANTLY GREATER FOR UNIFORM MAPS, U<S MEANS THAT THEY ARE SIGNIFICANTLY SMALLER. P-VALUES ARE GIVEN IN PARENTHESES.

All these factors suggest that uniform maps are quicker learners because they are better at generalising knowledge. This is more apparent for regular paradigms because they are less sensitive to token frequency effects and easier to generalise due to both their internal formal consistency (stem invariance) and their greater similarity to other paradigms (wordlikeness). To test the capacity of TSOMs to make generalisations, we checked their accuracy in recalling 50 unknown members of known paradigms in the two training regimes. Accuracy scores by learning epochs are reported in Figure 11, showing a significantly better performance by uniform maps, particularly in the 12-20 epoch range.

As an example of the interaction between frequency, regularity and acquisition, the plots in Figure 12 depict the learning time course of three German (sub)paradigms (*brauchen* ‘need’, *wollen* ‘want’ and *sein* ‘be’) exhibiting different degrees of inflectional (ir)regularity. In the regular *brauchen*, a uniform distribution (white circles in the left panel of Figure 12) appears to favour (i) earlier acquisition epochs than a skewed distribution does (black circles), and (ii) a short (2 to 3 epochs) acquisition span. When it comes to an irregular paradigm like *sein*, the pattern is

reversed. Since stems may vary across paradigm cells, acquisition is controlled by rote memorisation of individual forms. As a result, we observe a statistically significant ($p < .05$) advantage for acquisition in the skewed training regime (black circles in the right panel of Figure 12) than in a uniform training regime (white circles, same panel), a longer acquisition span than the one of *brauchen*, and a strong correlation ($r = -.825$, $p < .05$) between learning time and stem-alternant cumulative frequencies. In fact, *sein* offers an example of a low-entropy irregular paradigm, where suppletive stems are found in high-frequency forms.

The irregular paradigm of *wollen* (Figure 12, middle panel) represents an intermediate case between the previous two, where a relatively isolated stem like *will* can use its higher frequency to offset the prevalence of *woll*. As a result, there is no significant effect of either type or token frequency distributions.

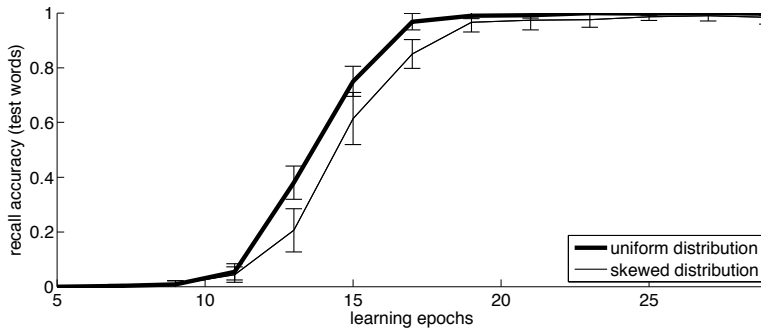


FIGURE 11. THE TIME COURSE OF WORD GENERALISATION BY UNIFORM MAPS (THICK LINE) AND SKEWED MAPS (THIN LINE) MAPS.

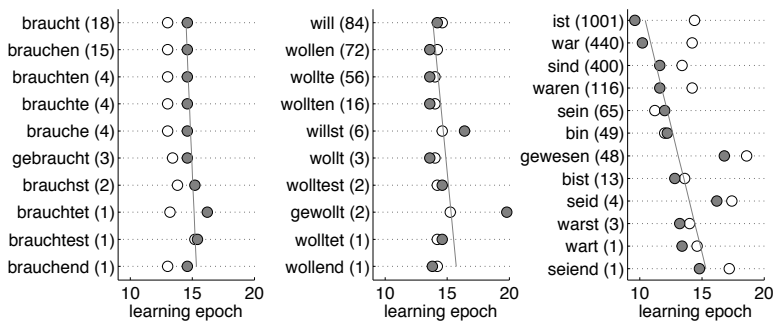


FIGURE 12. LEARNING EPOCHS FOR *BRAUCHEN* (LEFT), *WOLLEN* (MIDDLE) AND *SEIN* (RIGHT) WITH SKEWED (BLACK CIRCLES) AND UNIFORM (WHITE CIRCLES) DISTRIBUTIONS.

5. GENERAL DISCUSSION AND CONCLUDING REMARKS

Frequency reverberates on all levels of lexical organisation. There are frequency effects for individual words (TOKEN FREQUENCY EFFECTS), as well as frequency effects above the word level (cumulative TOKEN FREQUENCY EFFECTS for word families such as inflectional paradigms), and below the word level (both intra- and inter-paradigmatic TYPE FREQUENCY EFFECTS yielding sublexical patterns). This defines a hierarchy of frequency effects, which has far reaching consequences on the time course of lexical acquisition.

It should be appreciated that what favours acquisition on one hierarchical level can be a detrimental factor on another level. This is due to the effects of frequency on memory entrenchment, and to the effects of entrenchment on the emergence of relations between stored word forms.

High-frequency words are learned more quickly, and this is also true of high-frequency paradigms in general. However, regular paradigms are less affected by token frequency, as they can rely on the cumulative boost of type frequency. In addition, a uniform training condition tends to speed up acquisition of paradigms, with the only exception of a handful of high-frequency highly irregular paradigms, which are learned significantly faster in the skewed condition. Most of these are memory effects, reflecting the well-known correlation between (high) frequency and entrenchment. But some of them are not. For example, the across-the-board acquisitional boost provided by uniform training to regular as well as irregular paradigms is more reminiscent of a processing effect than a memory by-product.

Our experimental results show that in a low-entropy inflectional paradigm, high-frequency forms are learned at early epochs, followed by a few low-frequency forms that can benefit from early-acquired forms by “parasitically” exploiting the activation chains developed by their high-frequency paradigmatic counterparts. This appears to speed up the rate of acquisition of high-frequency and some low-frequency words of the same paradigm at early epochs, but it eventually delays acquisition of the entire paradigm, since early entrenchment and loss of plasticity disfavour later generalisations.

Generally speaking, high-frequency forms tend to be isolated both within their own paradigm and from other paradigms, and this keeps their morphological information from being readily spread. In a TSOM, this has a strong impact on developing biased expectations. INTRA-PARADIGMATICALLY, high-frequency words develop entrenched, tightly connected node chains, building up a strong expectation for few endings only. Other alternative endings are thus disfavoured, and the propensity of the map to acquire both low-frequency and novel endings goes down. INTER-PARADIGMATICALLY, deeply entrenched node chains share very little of their information with

other chains, thus preventing words of other paradigms from taking advantage of this information through shared connections. Conversely, uniform frequency distributions over members of the same inflectional paradigm (highly entropic paradigms) make members more readily accessible. Thus, less biased expectations favour global (PARADIGM-BASED) acquisition.

All of this makes an important connection with issues of perception of sublexical structure and morphological parsability. Entrenched lexical representations resist being perceived as morphologically complex. We contend that perception of morphological structure has to do with information sharing and high co-activation levels between word forms. A suffix is parsed/perceived as an independent element when many words share it, and this happens whenever many low-weighted, pre-synaptic connections lead to the same suffix, i.e. when many competing chains share it. On the contrary, a deeply entrenched chain is poorly interconnected with other chains and hardly co-activates them. High entropy levels over weighted pre-synaptic connections are thus conducive to perception of structure and to the emergence of morphological relations, and, ultimately, to acquisition of inflectional paradigms.

Symbolic and sub-symbolic views on morphological structure are classically implemented as combinatorial, rule-based models (e.g. Pinker & Prince 1988) and probabilistic associative models (e.g. Rumelhart & McClelland 1986) respectively. Morphology is viewed by the former as a combinatorial system of entries and rules, where regular word forms (unlike irregulars) are not memorised in the lexicon, but produced on-line. For the latter models, morphological structure is conceptualised as a non-discrete, epiphenomenal by-product of input-output mapping, irrespective of degrees of morphological regularity. In section 4, we examined the contrasting predictions made by the two accounts about the role of frequency and formal resemblance on the acquisition of regular vs. irregular inflectional paradigms, with each account predicting only part of the psycholinguistic evidence available.

But another view appears to suggest itself. There is converging evidence of a graded, highly distributed view of MORPHOLOGICAL STRUCTURE as an emergent property of lexical self-organisation. The view assumes that all word forms are memorised in the lexicon, thus making no distinction between a uniquely stored base form and all other non-base forms, which are processed on-line, in both recognition and production (see Baayen 2007, for an overview). In addition, to capture the fact that words encountered frequently exhibit different lexical properties from words encountered relatively infrequently, any model of the mental lexicon must assume that accessing a word in some way affects the access representation of that word (Foster 1976; Marslen-Wilson 1993; Sandra 1994; among others).

We showed that the tight connection between frequency, parsability and acquisition can be accounted for in terms of a self-organising model of serial cognition that blurs the distinction between memory (representations) and processing (operations defined over representations). Although our experimental results are grounded on storage of full forms, the emergent morphological organisation crucially relies on bound morphological constituents (stems and affixes), as opposed to full forms, and presupposes a level of organisation of memorised words into probabilistically connected sublexical parts. This seems to us a necessary step to take in analysing those languages whose morphology (unlike English morphology) is stem-based rather than word-based. In addition, it is an interesting by-product of our analysis that the explanatory notions we developed (e.g. entrenchment, information sharing, expectation and co-activation) meet recent psycholinguistic evidence on the role of frequency distributions in the perception of lexical relatedness by human speakers (Hay & Baayen 2002, 2005; Moscoso et al. 2004b).

The lexical structures developed by temporal self-organising maps neither amount to a full (hierarchical) listing of stored word forms (of the sort typically provided by a tree-like structure, as sketched in Figure 2, top), nor define redundancy-free, maximally compressed lexical representations (such as word graphs, Figure 2, middle). Each node is fully connected with any other node of the map. At the outset, weights are randomly distributed over connections. After training, the final amount of interconnectivity (defined by the adjustment of weights on each connection) is the graded result of the probabilistic support received by the map from a training set of unevenly distributed word forms, which exhibit different degrees of morphological redundancy. As connection weights can take any real value in the 0-1 interval, perception of structure at morpheme boundaries is graded by definition, ranging from no connection to fully dedicated connections. In this perspective, the overall structure of a self-organised lexicon is the end result of a graded composition (of the sort depicted in Figure 2, bottom) of tree-like and graph-like structures, which are taken to be somewhat limiting cases in the perception of lexical items.

An important consequence of this notion of graded structure is that it makes the distinction between memorisation and processing a matter of degree. Irregulars are known to be learned by rote because the process of storing them can hardly take advantage of knowledge of already stored words. This makes storage of irregulars acutely sensitive to token frequency effects. Memorising regulars, on the other hand, is often the end result of using past knowledge of already stored items and can thus be conceptualised as the outcome of a generalisation step, based on structure sharing and co-activation. Storing an item by generalisation is not sensitive to its token frequency but it has a distinctive processing quality: it depends

on the local propensity of the item to be generalised (its wordlikeness); on the global propensity of the map for spreading information through a global network of associative connections; on the amount of redundancy exhibited by the paradigm which the item belongs to; and it is a relatively instantaneous process. It is an interesting result of the present investigation that such a broad range of evidence, including the role of morphological structure in lexical organisation and the principled difference between regular and irregular morphological families in speakers' perception, can effectively be accounted for by a distributed, memory-based computational architecture.

In the end, the experimental evidence offered in these pages provides computational and empirical support to the hypothesis that emergence of morphological structure is set on by relevant formal redundancy in the lexicon, through processing-based perception of (sub)regular formatives shared by concurrently and competitively stored words. In this perspective, morphological structure emerges in a gradient fashion from associative relations among fully-inflected words, and lexical perception and organisation are grounded on memory-based processing strategies, where many input factors – such as word frequency, paradigm (ir)regularity, as well as salience, wordlikeness and word length – dynamically affect lexical acquisition and the development of morphological knowledge.

REFERENCES

- Alegre, M. & P. Gordon (1999). Frequency effects and the representational status of regular inflections. *Journal of Memory and Language* 40. 41-61.
- Baayen, H., R. Piepenbrock & L. Gulikers (1995). The CELEX lexical database (CD-ROM). Philadelphia: Linguistic Data Consortium.
- Baayen, R. H., R. Dijkstra & R. Schreuder (1997). Singulars and plurals in Dutch. Evidence for a parallel dual route model. *Journal of Memory and Language* 37. 94-117.
- Baayen, R. H., H. L. Wurm & J. Aycocock (2007). Lexical dynamics for low-frequency complex words. A regression study across tasks and modalities. *The Mental Lexicon* 2(3). 419-463.
- Baayen, R. H. (2007). Storage and computation in the mental lexicon. In G. Jarema & G. Libben (eds.), *The mental lexicon: Core perspectives*, 81-104. Amsterdam: Elsevier.
- Baddeley, A. D. (2007). *Working memory, thought and action*. Oxford: Oxford University Press.
- Bailey T. M. & U. Hahn (2001). Determinants of wordlikeness: Phonotactics or lexical neighborhoods? *Journal of Memory and Language* 44. 568-591.
- Bates E. & J. C. Goodman (1997). On the inseparability of grammar and the lexicon: Evidence from acquisition, aphasia and real-time processing. *Language and Cognitive Processes* 12. 507-584.

- Bates, E., & J. C. Goodman (1999). On the emergence of grammar from the lexicon. In B. MacWhinney (ed.), *The emergence of language*, 29-79. Mahwah, NJ: Lawrence Erlbaum.
- Burzio, L. (2004). Paradigmatic and syntagmatic relations in Italian verbal inflection. In J. Auger, J. C. Clements, & B. Vance (eds.), *Contemporary approaches to Romance linguistics*, 17-44. Amsterdam & Philadelphia: John Benjamins.
- Burani, C. & A. Laudanna (1992). Units of representation of derived words in the lexicon. In R. Frost & L. Katz (eds.), *Orthography, phonology, morphology and meaning*, 361-376. Amsterdam: Elsevier.
- Bybee, J. (1985). *Morphology. A study of the relation between meaning and form*. Amsterdam & Philadelphia: John Benjamins.
- Bybee, J. (1995). Regular morphology and the lexicon. *Language and Cognitive Processes* 10(5). 425-455.
- Bybee, J. (2002). Sequentiality as the basis of constituent structure. In T. Givón & B. Malle (eds.), *The evolution of language out of pre-language*, 107-132. Amsterdam & Philadelphia: John Benjamins.
- Catani, M., D. K. Jones & D. H. ffytche (2005). Perisylvian language networks of the human brain. *Annals of Neurology* 57. 8-16.
- Corbett, G., A. Hippisley & D. Brown (2001). Frequency, regularity and the paradigm: A perspective from Russian on a complex relation. In J. Bybee & P. Hopper (eds.), *Frequency and the emergence of linguistic structure*, 201-226. Amsterdam & Philadelphia: John Benjamins.
- Cowan, N. (2001). The magical number 4 in short-term memory: A reconsideration of mental storage capacity. *Behavioral and Brain Sciences* 24. 87-185.
- Duñabeitia, J. A., M. Perea & M. Carreiras (2008). Does darkness lead to happiness? Masked suffix priming effects. *Language and Cognitive Processes* 23. 1002-1020.
- Ellis, A. & C. Morrison (1998). Real age-of-acquisition effects in lexical retrieval. *Journal of Experimental Psychology: Learning, Memory and Cognition* 24. 515-523.
- Ellis, N. C. & R. Schmidt (1998). Rules or associations in the acquisition of morphology? The frequency by regularity interaction in human and PDP learning of morphosyntax. *Language and Cognitive Processes* 13. 307-336.
- Ferro, M., G. Pezzulo & V. Pirrelli (2010). Morphology, memory and the mental lexicon. In Pirrelli, V. (ed.), *Lingue e Linguaggio* IX(2). 199-238.
- Ferro, M., C. Marzi & V. Pirrelli (2011). A self-organizing model of word storage and processing: Implications for morphology learning. In *Lingue e Linguaggio* X(2). 209-226.
- Forster, K. I. (1976). Accessing the mental lexicon. In R. J. Wales & E. Walker (eds.), *New approaches to language mechanisms*, 257-287. Amsterdam: North-Holland.
- Gathercole, S. E. & A. D. Baddeley (1989). Evaluation of the role of phonological STM in the development of vocabulary in children: A longitudinal study. *Journal of memory and language* 28(2). 200-213.
- Hay, J. B. (2001). Lexical frequency in morphology: is everything relative? *Linguistics* 39. 1041-1070.

- Hay, J. B. & H. R. Baayen (2002). Parsing and productivity. In Booij, G. & J. van Marle, *Yearbook of Morphology 2001*, 203-235. Dordrecht: Kluwer Academic Publishers.
- Hay, J. B. & I. Plag (2004). What constrains possible suffix combinations? On the interaction of grammatical and processing restrictions in derivational morphology. *Natural Language & Linguistic Theory* 22. 565-596.
- Hay, J. B. & H. R. Baayen (2005). Shifting paradigms: Gradient structure in morphology. *Trends in Cognitive Sciences* 9. 342-348.
- Kohonen, T. (2001). *Self-organizing maps*. Berlin-Heidelberg: Springer Verlag.
- Koutnik, J. (2007). Inductive modelling of temporal sequences by means of self-organization. In *Proceeding of International Workshop on Inductive Modelling*, Prague, 269-277.
- Marslen-Wilson, W. (1993). Issues of representation in lexical access. In *Cognitive models of speech processing: The Second Sperlonga Meeting*, 187-210.
- Marslen-Wilson, W., M. Ford, L. Older & X. Zhou (1996). The combinatorial lexicon: priming derivational affixes. In G. Cotrell (ed.), *Proceedings of the 18th Annual Conference of the Cognitive Science Society*, 223-227. Mahwah, NJ: Lawrence Erlbaum.
- Marzi C., M. Ferro & V. Pirrelli (2012a). Prediction and generalisation in word processing and storage. In *On-line Proceedings of the 8th Mediterranean Morphology Meeting*, 113-130.
- Marzi C., M. Ferro & V. Pirrelli (2012b). Word alignment and paradigm induction. *Lingue e Linguaggio* XI (2). 251-274.
- Marzi C., M. Ferro, C. Caudai & V. Pirrelli (2012c). Evaluating Hebbian self-organizing memories for lexical representation and access. In *Proceedings of 8th International Conference on Language Resources and Evaluation (ELRA – LREC 2012)*. 886-893.
- Moscoso del Prado Martín, F., A. Kostić & H. Baayen (2004a). Putting the bits together: an information theoretical perspective on morphological processing. *Cognition* 94. 1-18.
- Moscoso del Prado Martín, F., R. Bertram, T. Häikiö, R. Schreuder & H. Baayen (2004b). Morphological family size in a morphologically rich language: The case of Finnish compared with Dutch and Hebrew. *Journal of Experimental Psychology: Learning, Memory and Cognition* 30(6). 1271-1278.
- Papagno, C., T. Valentine & A.D. Baddeley (1991). Phonological short-term memory and foreign-language vocabulary learning. *Journal of Memory and Language* 30. 331-347.
- Pinker, S. & A. Prince (1988). On language and connectionism: Analysis of a parallel distributed processing model of language acquisition. *Cognition* 29. 195-247.
- Pinker, S. (1999). *Words and rules: The ingredients of language*. New York: Harper Collins.
- Pirrelli, V., M. Ferro & B. Calderone (2011). Learning paradigms in time and space. Computational evidence from Romance languages. In M. Goldbach, M. O. Hinzelin, M. Maiden & J. C. Smith (eds.), *Morphological autonomy: Perspectives from Romance inflectional morphology*, 135-157. Oxford: Oxford University Press.

- Plunkett, K. & V. Marchman (1993). From rote learning to system building – acquiring verb morphology in children and connectionist nets. *Cognition* 48. 21-69.
- Rumelhart, D. & J. McClelland (1986). On learning the past tense of English verbs. In D. E. Rumelhart & J. McClelland (eds.), *Parallel distributed processing: Explanations in the microstructure of cognition*, 216-271. Cambridge, MA: The MIT Press.
- Sandra, D. (1994). The morphology of the mental lexicon: Internal word structure viewed from a psycholinguistic perspective. *Language and Cognitive Processes*. 9(3). 227-269.
- Smolík, F. (2010). Inflectional suffix priming in Czech verbs and nouns. In S. Ohlsson & R. Catrambone (eds.), *Proceedings of the 32nd Annual Conference of the Cognitive Science Society*, 1667-1672. Austin, TX: Cognitive Science Society.
- Taft, M. & K. I. Forster (1975). Lexical storage and retrieval of prefixed words. *Journal of Verbal Learning and Verbal Behavior* 14. 638-647.
- Whaley, C. P. (1978). Word-nonword classification time. *Journal of Verbal Learning and Verbal Behavior* 17. 143-154.
- Zevin, J. D. & M. S. Seidenberg (2002). Age-of-acquisition effects in reading and other tasks. *Journal of Memory and Language* 47. 1-29.

Claudia Marzi

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

Marcello Ferro

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

Vito Pirrelli

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