Lexical emergentism and the "frequency-by-regularity" interaction

Claudia Marzi **Marcello Ferro** **Vito Pirrelli**
Institute for Computational Linguistics - National Research Council - Pisa
{claudia.marzi,marcello.ferro,vito.pirrelli}@ilc.cnr.it

Abstract

In spite of considerable converging evidence of the role of inflectional paradigms in word acquisition and processing, little efforts have been put so far into providing detailed, algorithmic models of the interaction between lexical token frequency, paradigm frequency, paradigm regularity. We propose a neuro-computational account of this interaction, and discuss some theoretical implications of preliminary experimental results.

1 Introduction

Over the last fifteen years, growing evidence has accrued of the role of morphological paradigms in the developmental course of word acquisition. Children have been shown to be sensitive to sub-regularities holding among paradigm cells (see, among others, Orsolini et al., 1998; Laudanna et al., 2004 on Italian; Dabrowska, 2004, 2005 on Polish; and Labelle and Morris, 2011 on French). In line with this evidence, and contrary to both rule-based (e.g. Pinker and Ullman, 2002; Albright, 2002) and connectionist approaches to word acquisition (Rumelhart and McClelland, 1986), no unique paradigm cell can be identified as the base source of all inflected forms produced by the speaker, but the structure of the entire paradigm is understood to play a fundamental role in both word acquisition and processing.

Such evidence supports a view of the mental lexicon as an emergent integrative system, whereby words are concurrently, redundantly and competitively stored (Alegre and Gordon, 1999; Baayen et al., 2007). The view assumes that all word forms are memorised in the lexicon, thus making no distinction between regular and irregular inflected forms, or between uniquely stored bases and all other non-base forms produced by the speaker on demand (see Baayen, 2007; Marzi, 2014; for a recent overview). In addition, to capture the fact that words encountered frequently exhibit different lexical properties from words encountered relatively infrequently, any model of lexical access must assume that accessing a word in some way affects the access representation of that word (e.g. Foster, 1976; Marslen-Wilson, 1993; Sandra, 1994).

In spite of such a wealth of converging evidence, however, little efforts have been put so far into providing detailed, algorithmic models of the interaction between word frequency, paradigm frequency, paradigm regularity and lexical familiarity in word acquisition and processing. We offer here such an algorithmic account, and discuss some theoretical implications on the basis of computational simulations.

2 The computational model

In the present contribution, we use Temporal Self-organising Maps (TSOMs) to simulate dynamic effects of lexical storage, organisation and competition.

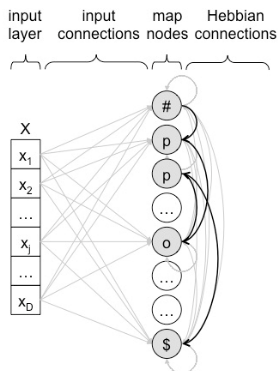

Figure 1. An integrated activation pattern for the input string "*#pop$*". Note that two distinct, but topologically neighbouring nodes respond to the two *p*'s in *pop*, bearing witness to the process of selective sensitivity to time-bound instances of the same symbol type. For simplicity, only the nodes that are most highly activated by each input symbol are shaded and tagged with that symbol.

TSOMs, a variant of classical Kohonen's SOMs (Kohonen, 2001), are dynamic memories that are trained to store and classify time-series of symbols through patterns of activation of fully interconnected nodes (Koutnik, 2007; Ferro et al., 2010; Pirrelli et al., 2011; Marzi et al., 2012). Map nodes mimic neural clusters, with inter-node connections representing neuron synapses whose weights determine the amount of influence that the activation of one node has on another node (Fig. 1). Each map node receives input

In Vito Pirrelli, Claudia Marzi, Marcello Ferro (eds.): *Word Structure and Word Usage*. Proceedings of the NetWordS Final Conference, Pisa, March 30-April 1, 2015, published at http://ceur-ws.org

connections from an input layer where individual symbols making up a word are presented one at a time, in their order of appearance. Input connections thus convey information of the current input stimulus to map nodes. Hebbian connections, on the other hand, are strengthened each time two nodes are activated at consecutive time ticks, conveying the probabilistic expectation that one node will be activated soon after another node is activated.

When a symbol is shown on the input layer at a certain time tick, all map nodes are fired synchronously, their overall pattern of activation representing the processing response of a TSOM to the symbol at that time tick. Due to principles of topological organisation of map's responses, similar input stimuli (i.e. two instances of the same symbol in different contexts) tend to be associated with largely overlapping memory traces (e.g. the two *p* nodes activated by *pop* in Fig. 1). During training, nodes get gradually specialised to respond most strongly to specific time-bound instantiations of symbols, while remaining relatively inactive in the presence of other stimuli. A recurrent activation pattern associated with an input symbol occurring in a specific context can thus be seen as the map's memory trace for that symbol in that context.

An input word is administered to a TSOM as a time series of symbols, i.e. a sequence of letters or sounds presented on the input layer one at a time. The map's response to a word stimulus is the overall activation pattern obtained through integration of the activation patterns triggered by the individual symbols making up the word (see Fig. 1 for a simplified example with the word *pop*). Accordingly, if two input strings present some symbols in common (e.g. *pop* and *cop*, *write* and *written*), they will tend to activate largely overlapping patterns of strongly responsive nodes. Like in the case of individual symbols, the integrated activation pattern for an input word is, at the same time, the systematic processing response of the map to an input stimulus, and the word's memorised representation (or memory trace) in the map.

To investigate issues of "frequency-by-regularity" interaction (Ellis and Smith, 1998), we compared two sets of parallel experiments carried out on German verb paradigms (Marzi et al., 2014) and Italian verb paradigms. By keeping constant some input conditions, such as selection of paradigm cells and degrees of morphological redundancy within training paradigms, while varying others, such as the frequency distribution of paradigm members, we can investigate the relative contribution of input factors to the timing and pace of lexical acquisition and suggest an explanatory account of their interaction.

3 Experimental evidence

Fifty German and fifty Italian verb (sub)paradigms were selected among the most highly ranked paradigms by cumulative frequency in a reference corpus (CELEX Lexical database for German, Baayen et al., 1995; Paisà Corpus for Italian, Lyding et al., 2014). For each paradigm, an identical set of 15 cells was used for training, for an overall number of 750 inflected forms for each language. Each data set was administered to the map for 100 epochs under two different training regimes: a uniform distribution (UD: 5 tokens per word), and a function of real word frequency distributions in the reference corpus (SD: tokens are in the range of 1 to 1000). By varying frequency and comparing the inflectional complexity of training data across the two experiments, we expected to gain some insights into the interplay between morphological regularity (defined by levels of predictability in stem and ending allomorphy of training data in the two languages) and word frequency in word acquisition. After training, we monitored the behaviour of the four resulting TSOMs (namely UD Italian, SD Italian, UD German and SD German) by controlling the time of acquisition of individual words, the time of acquisition of entire paradigms, and their acquisitional time span. For our present purposes, we define the time of acquisition of a single word as the training epoch whence a TSOM can accurately recall the word in question from its memory trace. Recall is a difficult task that requires that the map has developed a clear notion of how to unfold a synchronous activation pattern (the word's memory trace) into a sequence of nodes representing the correct letters making up the word, in the appropriate order. Likewise, for each paradigm, its time of acquisition by a map is the mean acquisition epoch of all forms belonging to the paradigm.

As a general trend, TSOMs acquire word forms by token frequency, with higher-frequency words being successfully recalled at earlier learning epochs. However, when it comes to the actual timing of paradigm acquisition, things get considerably more complex, with the notion of morphological regularity interacting non-trivially with token frequency distributions. In fact, in both

German and Italian, the vast majority of paradigms are acquired earlier (p<.005) in a UD regime than in an SD regime (Fig. 2).

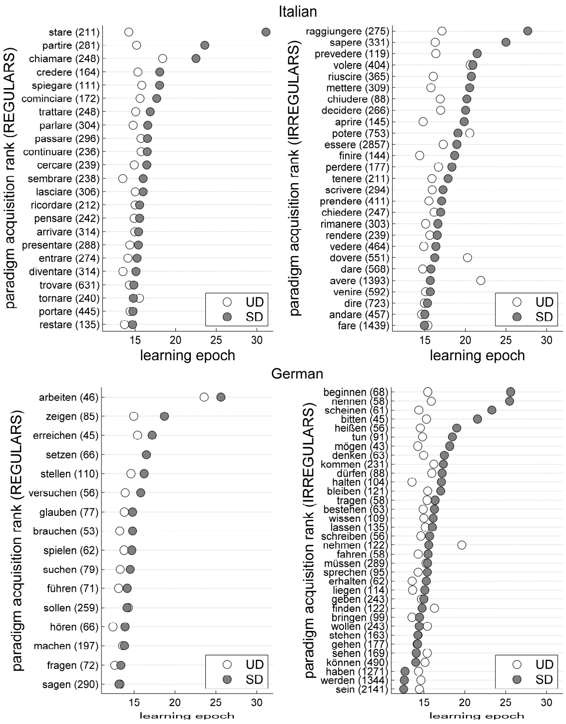

Figure 2: Time course of regular (left) and irregular (right) paradigms ranked by increasing learning epoch under SD (grey circles) and UD (white circles) regimes for both Italian (top) and German (bottom). Values are averaged across 5 map instances for each type.

4 Frequency by regularity interaction

Our simulations show that, in both languages, word forms in regular paradigms tend to be acquired earlier (significantly earlier learning epochs, p<.001), and regular paradigms are acquired more quickly (significantly shorter learning spans, i.e. lower number of epochs between the acquisition time of the first and the last member of a paradigm, p<.005) than irregular paradigms are. In German data, regular paradigms are less sensitive to token frequency effects than irregular paradigms are, as witnessed by the strong correlation (r=.95, p<.00001) between the time course of acquisition of regular paradigms in SD and UD regimes (Fig. 2, bottom left panel). Token frequency affects the acquisition of regular paradigms to a lesser extent than the acquisition of irregular ones, because regular stems can take advantage of their cumulative frequency across the whole paradigm. In fact, forms in regular paradigms exhibit a significant correlation between stem cumulative frequency and time of acquisition (r=-.40, p<.00001). Similarly, also German irregular paradigms, which exhibit a predictable stem allomorphy due to a limited number of alternants, show a correlation between stem cumulative frequency and acquisition time (r=-.24 p<.00001).

Conversely, in Italian, where verb conjugation exhibits more extensive and less predictable patterns of allomorphy than in German (Pirrelli, 2000), acquisition of irregular paradigms does not appear to benefit from stem cumulative token frequencies (r=.01, p>.5). This suggests that extensive allomorphy in a paradigm tends to minimise the influence of cumulative frequency on its acquisition, and isolated forms can only take advantage of their own token frequency, while taking no advantage of the frequency boost provided by other cells of the same paradigm. As a result, Italian irregular paradigms are acquired significantly (p<.005) later than their German homologues.

Our data cannot be explained away as a simple by-product of word-frequency effects. Experiments provide, in fact, evidence of interactive processing effects in word acquisition, whereby morphological regularity modulates frequency. Data analysis shows that recurrent patterns appear to determine global co-organisation of stored word forms and distributed, overlapping memory traces, which ultimately favour generalisation in lexical acquisition. Forms containing recurrent patterns can take advantage of the memory traces shared with other related forms, namely forms sharing the same stem, and connections between the nodes making up their memory traces are strengthened since patterns are shown more often in training, similarly to high-frequency isolated words.

This is particularly true for regular, highly entropic paradigms, i.e. those regular paradigms whose members exhibit uniform frequency distributions, and for irregular highly systematic paradigms. Conversely, where memory traces overlap less systematically, this effect is considerably reduced, as witnessed by the difference in time of acquisition between regular and irregular paradigms, particularly in Italian conjugation.

In TSOMs, the effects are the dynamic result of two interacting dimensions of memory self-organisation: (i) the syntagmatic or linear dimension, which controls the level of predictability and entrenchment of memory traces in the lexicon through the probabilistic distribution of weights over inter-node Hebbian connections; and (ii) the paradigmatic or vertical dimension, which controls for the number of similar, paradigmatically-related word forms that get co-activated when one member of a paradigm is input to the map (Pirrelli et al., 2014).

High-frequency words develop quick entrenchment of Hebbian connections, which eventually cause high levels of node activation in their memory traces and sparser co-activation of memory traces of other words. Strong connections and high activation levels mean high expectations for frequently activated memory traces, which are thus recalled more easily and are less confusable with other neighbouring words. Likewise, in regular and sub-regular paradigms, sharing memory traces can strengthen connections and raise node activation levels, since all related forms can take advantage of the memory traces shared with other members of the same paradigm.

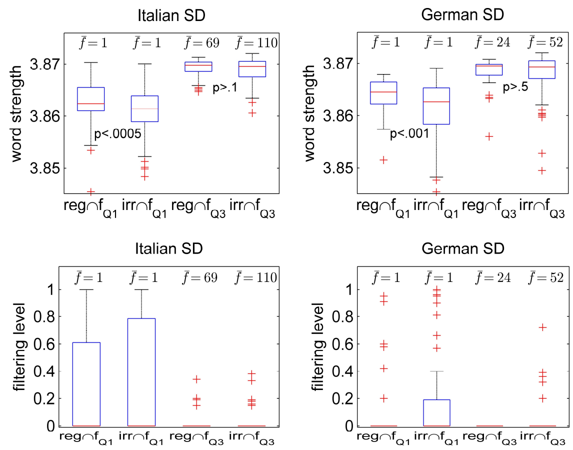

Figure 3: Levels of activation strength (top) and filtering (bottom) for Italian (left) and German (right), for four regularity-by-frequency classes. Low-frequency is set below the first quartile of frequency distributions in the two training sets, while high-frequency being set above the third quartile.

This dynamic provides an algorithmic account of the observation that regularity favours acquisition of both high- and low-frequency words, as shown in Fig. 3, where we compare average levels of activation for four classes of training word forms: low-frequency regulars, low frequency irregulars, high-frequency regulars and high-frequency irregulars.[1]

Activation levels of low-frequency words appear to be significantly stronger within regular paradigms than within irregular paradigms (Fig. 3, top). Stronger activation levels make patterns less confusable and easier to be accessed, as witnessed by the lower level of filtering[2] required for activation patterns to be recalled accurately (Fig. 3, bottom). We observe, in fact, a highly significant correlation ($r=.49$, $p<.00001$ for both datasets) between levels of filtering and words' learning epochs.

High-frequency words predictably show higher activation levels than low-frequency words, with an interesting difference of the interaction of frequency and activation levels of regulars and irregulars. High-frequency, highly irregular words (e.g. German *ist* or Italian *è*) are stored in isolation, with highly-activated memory nodes and no co-activation with other words. As a result, they require little filtering to be recalled and are acquired considerably quickly. High-frequency regular paradigms, despite in both Italian and German training sets their average frequency is nearly half the average frequency of high-frequency irregulars, show comparable levels of activation with high-frequency irregulars, due to the facilitatory effect of having more words that consistently activate the same pattern of nodes.

This evidence shows that regularity indeed modulates the interaction between frequency and activation strength, and it gives a strong indication that acquisition of regulars is typically paradigm-based, whereas acquisition of irregulars is mostly item-based.

Surely, as the notion of paradigm regularity is inherently graded, some verb systems show higher sensitivity to these effects than others. This is illustrated by German sub-regular paradigms, which present fewer and more predictable stem alternants than Italian sub-paradigms, and thus larger stem-sharing word families. Accordingly, TSOMs allocate comparatively higher levels of activation to low-frequency German sub-regulars and acquire them earlier than their Italian homologues.

The evidence reported here establishes, in our view, an important connection between aspects of morphological structure, frequency distributions of words in paradigms, and lexical acquisition in concurrent, competitive storage. Acquisition of redundant morphological patterns play an increasingly important role in an emergent lexicon, shifting acquisitional strategies from rote memorisation (typical of irregular low-entropy paradigms) to dynamic memory-based generalisation.

[1] Frequency thresholds are set below the first quartile (low frequency) and above the third quartile (high frequency) in the frequency distribution of training word forms.

[2] Filtering an integrated activation pattern refers to the process of bringing down to zero the levels of activation of nodes that do not reach a set threshold.

References

Maria Alegre and Peter Gordon. 1999. Frequency effects and the representational status of regular inflections. *Journal of Memory and Language*, 40: 41-61.

Harald R. Baayen, Richard Piepenbrock and Leon Gulikers. 1995. The CELEX Lexical Database (CD-ROM). Philadelphia: Linguistic Data Consortium.

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

Lucia Colombo, Alessandro Laudanna, Maria De Martino and Cristina Brivio. 2004. Regularity and/orconsistency in the production of the past participle? *Brain and Language*, 90: 128-142.

Ewa Dabrowska. 2004. Rules or schemata? Evidence from Polish. *Language and cognitive processes*, 19 (2): 225–271.

Ewa Dabrowska. 2005. Productivity and beyond: mastering the Polish genitive inflection. *Journal of child language*, 32: 191-205.

Nick C. Ellis and Richard 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.

Marcello Ferro, Giovanni Pezzulo and Vito Pirrelli. 2010. Morphology, Memory and the Mental Lexicon. In Pirrelli, V. (ed.), *Lingue e Linguaggio*, IX(2): 199-238.

Jan Koutnik. 2007. Inductive Modelling of Temporal Sequences by Means of Self-organization. In *Proceeding of International Workshop on Inductive Modelling*. Prague: 269-277.

Marie Labelle and Lori Morris. 2011. The acquisition of a verbal paradigm: Verb Morphology in French L1 children. *Prépublication.* (Montréal, Québec, Canada, UQAM, département de linguistique). http://www.archipel.uqam.ca/3992/1/Labelle-Morris_AcquisitionVerbalParadigm.pdf

Verena Lyding, Egon Stemle, Claudia Borghetti, Marco Brunello, Sara Castagnoli, Felice Dell'Orletta, Henrik Dittmann, Alessandro Lenci and Vito Pirrelli. 2014. The PAISÀ Corpus of Italian Web Texts. In F. Bildhauer and R. Schäfer (eds.) *Proceedings of the 9th Web as Corpus Workshop (WaC-9)*: 36-43. Gothenburg.

Claudia Marzi, Marcello Ferro and Vito Pirrelli. 2012. Word alignment and paradigm induction. *Lingue e Linguaggio*, XI (2): 251-274.

Claudia Marzi, Marcello Ferro and Vito Pirrelli. 2014. Morphological structure through lexical parsability. *Lingue e Linguaggio*, XIII (2): 263-290.

Claudia Marzi. 2014. Models and dynamics of the morphological lexicon in mono- and bilingual acquisition. *Unpublished PhD Dissertation.* University of Pavia. www.comphyslab.it/redirect/?id=claudia.marzi.en_phd

Margherita Orsolini, Rachele Fanari and Hugo Bowles. 1998. Acquiring regular and irregular inflections in a language with verb classes. *Language and cognitive processes*, 13(4): 425-464.

Steven Pinker and Michael Ullman. 2002. The past and future of the past tense. *Trends in Cognitive Science*, 6: 456-463.

Vito Pirrelli, Claudia Marzi and Marcello Ferro. 2014. Two-dimensional Wordlikeness Effects in Lexical Organisation. In: Basili R., Lenci A., Magnini B. (eds.) *Proceedings of the First Italian Conference on Computational Linguistic*, December 9-11, 2014. 301-305, Pisa: Pisa University Press.

Vito Pirrelli, Marcello Ferro and Basilio Calderone. 2011. Learning paradigms in time and space. Computational evidence from Romance languages. In M. Maiden, J. C. Smith, M. Goldbach and M. O. Hinzelin (eds.), *Morphological Autonomy: Perspectives from Romance Inflectional Morphology*, 135-157. Oxford: Oxford University Press.

Vito Pirrelli. 2000. *Paradigmi in morfologia. Un approccio interdisciplinare alla flessione verbale dell'italiano*. Pisa-Roma: Istituti editoriali e poligrafici internazionali.

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

David E. Rumelhart and James L. McClelland. 1986. On learning the past tense of English verbs. In McClelland, J.L. and Rumelhart, D.E. (eds.) *Parallel distributed processing*, 217-270. Cambridge: MIT Press.