10.12871/CLICIT2014158

Two-dimensional Wordlikeness Effects in Lexical Organisation

Vito Pirrelli
ILC-CNR Pisa
vito.pirrelli@ilc.cnr.it

Claudia Marzi
ILC-CNR Pisa
claudia.marzi@ilc.cnr.it

Marcello Ferro
ILC-CNR Pisa
marcello.ferro@ilc.cnr.it

Abstract

English. The main focus of research on wordlikeness has been on how serial processing strategies affect perception of similarity and, ultimately, the global network of associative relations among words in the mental lexicon. Comparatively little effort has been put so far, however, into an analysis of the reverse relationship: namely, how global organisation effects influence the speakers' perception of word similarity and of words' internal structure. In this paper, we explore the relationship between the two dimensions of wordlikeness (the "syntagmatic" and the "paradigmatic" one), to suggest that the same set of principles of memory organisation can account for both dimensions.

Italiano. *Gran parte dei lavori sulla nozione di "familiarità lessicale" ha analizzato come le strategie di elaborazione seriale influenzino la percezione della similarità all'interno della rete di relazioni formali nel lessico mentale. Poca attenzione è stata tuttavia dedicata finora a come queste relazioni globali influenzino la percezione della similarità lessicale. L'articolo esplora questa interconnessione tra relazioni sintagmatiche e paradigmatiche, attribuendola a un insieme omogeneo di principi per l'organizzazione della memoria seriale.*

1 Introduction

The language faculty requires the fundamental ability to retain sequences of symbolic items, access them in recognition and production, find similarities and differences among them, and assess their degree of typicality (or WORDLIKENESS) with respect to other words in the lexicon. In particular, perception of formal redundancy appears to be a crucial precondition to morphology induction, epitomised by the so-called WORD ALIGNMENT problem. The problem arises whenever one has to identify recurrence of the same pattern at different positions in time, e.g. *book* in *handbook*, or *mach* in both German *macht* and *gemacht*. Clearly, no "conjunctive" letter coding scheme (e.g., Coltheart et al. 2001; Harm & Seidenberg 1999; McClelland & Rumelhart 1981; Perry et al. 2007; Plaut et al. 1996), which requires that the representation of each symbol in a string be anchored to its position, would account for such an ability. In Davis' (2010) SPATIAL ENCODING, the identity of the letter is described as a Gaussian activity function whose max value is centred on the letter's actual position, enforcing a form of fuzzy matching, common to other models disjunctively encoding a symbol and its position (Grainger & van Heuven 2003; Henson 1998; Page & Norris 1998, among others).

The role of specific within-word letter positions interacts with short-term LEXICAL BUFFERING and LEXICALITY effects. Recalling a stored representation requires that all symbols forming that representation are simultaneously activated and sustained in working memory, waiting to be serially retrieved. Buffering accounts for the comparative difficulty in recalling long words: more concurrently-activated nodes are easier to be confused, missed or jumbled than fewer nodes are. Notably, more frequent words are less likely to be confused than low-frequency words, since long-term entrenchment improves performance of immediate serial recall in working memory (Baddeley 1964; Gathercole et al. 1991).

Serial (or syntagmatic) accounts of local ordering effects in word processing are often complemented by evidence of another, more global (or paradigmatic) dimension of word perception, based on the observation that, in the normal course of processing a word, other non-target neighbouring words become active. In the word recognition literature, there is substantial agreement on the inhibitory role of lexical neighbours (Goldinger et al. 1989; Luce & Pisoni 1998; Luce et al. 1990). Other things being equal, target words with a large number of neighbours take more time to be recognised and repeated, as they suffer from their neighbours' competition in lexical buffering. This is particularly true when the target word is low-frequency. Nonetheless, there is contrasting evidence that dense neighbourhoods may speed up word reading time rather than delaying it (Huntsman & Lima 2002), and that high-entropy word families make their members more readily accessible than low-entropy families (Baayen et al. 2006).

Marzi et al. (2014) provide clear computational evidence of interactive effects of paradigm regularity and type/token lexical frequency on the acquisition of German verb inflection. Token frequency plays a paramount role in item-based learning, with highly frequent words being acquired at comparatively earlier stages than low-frequency words. Morphological regularity, on the other hand, has an impact on paradigm acquisition, regular paradigms being learned, on average, within a shorter time span than fully or partially irregular paradigms. Finally, frequency distribution of paradigmatically-related words significantly interacts with morphological regularity. Acquisition of regular paradigms depends less heavily on item-based storage and is thus less affected by differences in frequency distributions of paradigm members. Conversely, irregular paradigms are less prone to be generalised through information spreading and their acquisition mainly relies on itemised storage, thus being more strongly affected by the frequency distribution of paradigm members and by frequency-based competition, both intra- and inter-paradigmatically.

We suggest that compounded evidence of wordlikeness and paradigm frequency effects can be accounted for within a unitary computational model of lexical memory. We provide here preliminary evidence in this direction, by looking at the way a specific, neuro-biologically inspired computational model of lexical memories, Temporal Self-Organising Maps (TSOMs), accounts for such effects.

2 TSOMs

TSOMs are a variant of classical Kohonen's SOMs (Kohonen 2001), augmented with re-entrant Hebbian connections defined over a temporal layer encoding probabilistic expectations upon immediately upcoming stimuli (Koutnik 2007; Ferro et al. 2010; Pirrelli et al. 2011; Marzi et al. 2012a, 2012b). TSOMs consist of a network of memory nodes simultaneously responding to time-bound stimuli with varying levels of activation (Fig. 1). Through learning, nodes acquire selective sensitivity to input stimuli, i.e. they respond more strongly to a particular class of stimuli than to others. Selective sensitivity is based on both nature of the stimulus (through *what* connections), and stimulus timing (through *when* connections) (see Fig. 1). Accordingly, more nodes tend to be recruited to respond to the same symbol, each node being more sensitive to a specific occurrence of the symbol in context.

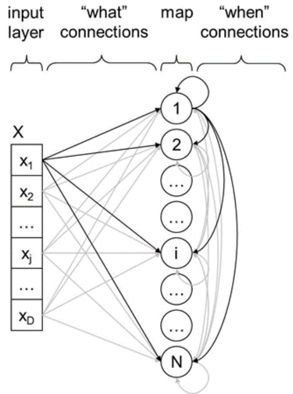

Figure 1: Outline architecture of a TSOM.

TSOMs can be trained on word forms as time-bound sequences of symbols by presenting each symbol on the input layer one at a time. The pattern of node activation prompted by each symbol is eventually integrated into a word memory trace, whose top-most activated nodes are named Best Matching Units (BMUs). Co-activation of the same BMUs by different input words reflects the extent to which the map perceives formal redundancies between words. We contend that perception of wordlikeness and morphological structure has to do with information sharing and co-activation levels between word forms.

2.1 Activation and co-activation effects

Two quantitative correlates have been suggested to account for effects of human perception of wordlikeness: N-GRAM PROBABILITY DENSITY (the likelihood that a word form results from concatenation of sublexical chunks of *n* length), and LEXICAL DENSITY (the number of word forms in the lexicon that are similar to a specific target word) (Bailey & Hahn 2001).

The two measures are highly correlated and thus easy to be confounded in measuring their independent effects on lexical tasks (Bard 1990). Bailey and Hahn (2001) propose to define *n*-gram probability densities in terms of the geometric mean of the product of the independent probabilities of bigram and trigram types extracted from the lexicon. In addition, following Luce and Pisoni (1998), the lexical neighbourhood of a target word can be defined as the set of word forms obtained from the target by substitution, deletion or insertion of one symbol.

With a view to establishing functional correlates between the behaviour of a TSOM and evidence of probability and neighbourhood density effects on word processing, we trained 10 instances of a TSOM on 700 uniformly-distributed Italian verb forms, belonging to a fixed set of 14 cells of the 50 most frequent verb paradigms in the Italian Tree Bank (Montemagni et al. 2003). We tested the 10 map instances on the task of RECALLING [1] four data sets: (i) the original TRAINING SET; (ii) a set of 50 TEST WORDS sampled from the same 50 paradigms of the original training set; (iii) an additional set of novel Italian verb paradigms which were not part of the original training (hereafter NOVEL WORDS); iv) a set of German verb forms (or Italian NON-WORDS).

On training and test words, accuracy is respectively 99.2% and 96.4%. On novel words, recall (44.4%) significantly correlates with both node ACTIVATION STRENGTH (r=0.471, p<.00001), i.e. the per-word mean activation level of BMUs in the words' memory traces, and between-node CONNECTION STRENGTH (r=0.506, p<.00001), i.e. the per-word mean strength of the temporal connections between consecutives BMUs. Equally significant but lower correlations with activation and connection strengths are found for recall scores on non-words (12.4%): r=0.335, p<.00001 and r=0.367, p<.00001. We observe that recall scores somewhat reflect a word familiarity gradient, ranging from known words (training set) and known word stems with novel inflections (test words) to novel paradigms (novel words) and non-words. In particular, the gradient reflects the extent to which a map has developed expectations on incoming words, which in turn are encoded as weights on temporal connections. Both connection and activation strength thus capture probabilistic knowledge of Italian orthotactic constraints.

In fact, n-gram probability does not explain recall scores entirely. Forward probabilities account for degrees of entrenchment in integrated memory traces but they say nothing about co-activation of other formally-related words. This information has to do with neighbourhood density and is controlled by the degree of global lexical co-activation by an input word, i.e. by the extent to which the word memory trace reverberates with all other memory traces in the lexicon (Fig. 4). Note that both test and novel words exhibit comparatively high levels of global co-activation, in contrast with non-words, whose degree of paradigmatic wordlikeness is consistently poorer (p<.00001).

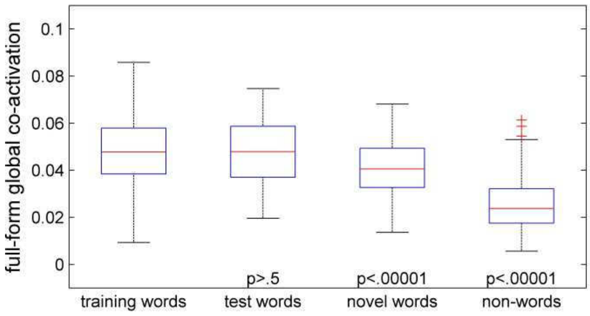

Figure 4: Per-word global co-activation.

We explain this overall effect by looking at differential values of activation strengths for stems and suffixes in Fig. 5. Here, Italian novel words score more highly on suffixes than on stems. As expected, they are recalled consistently more poorly, but their degree of perceived familiarity is due to their fitting Italian recurrent morphological patterns.

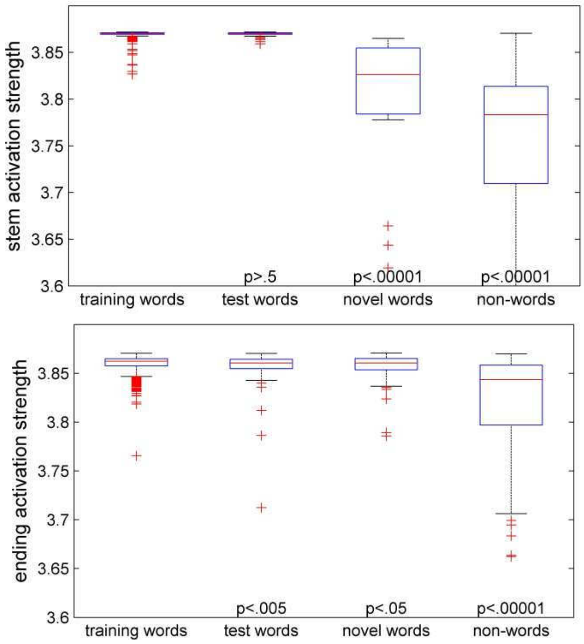

Figure 5: Per-stem (top panel) and per-ending (bottom panel) activation strength.

2.2 Frequency effects

In section 1, we overviewed contrasting evidence of inhibitory and facilitatory effects of neighbourhood density and neighbourhood frequency in different word processing tasks. To test Baayen and colleagues' claim that large, evenly-distributed word families facilitate accessibility of their own members, we assessed, for each Ital-

[1] Recall simulates the process of retrieving a sequence of letters from an integrated word memory trace. A word is recalled accurately if the map retrieves all its symbols in the correct left-to-right order.

ian word in our training set, the level of confusability of its memory trace on the map in the recall task. A word memory trace contains, over and above target BMUs, also nodes that are associated with concurrently activated neighbours. By increasingly filtering out nodes with lower activation levels in the trace, we can make it easier for the map to reinstate the appropriate sequence of target nodes by eliminating spurious competitors. Fig. 6 shows the box plot distribution of the mean filtering level for classes of words having up to 2 neighbours, between 3 and 12 neighbours, and more than 12 neighbours.

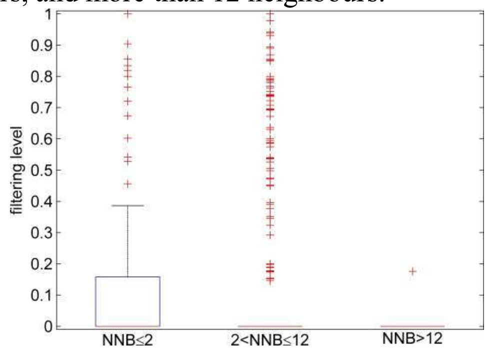

Figure 6: Filtering levels on word memory traces for serial recall, for three neighbourhood-density bins.

In TSOMs, words with sparser neighbours require more filtering to be recalled correctly from their memory traces. This is due to the facilitatory effect of having more words that consistently activate the same sequences of nodes. Fewer neighbours weaken this effect, making it more difficult to recover the right sequence of nodes from a word memory trace. This greater difficulty is reflected by larger filtering levels in Fig. 6.

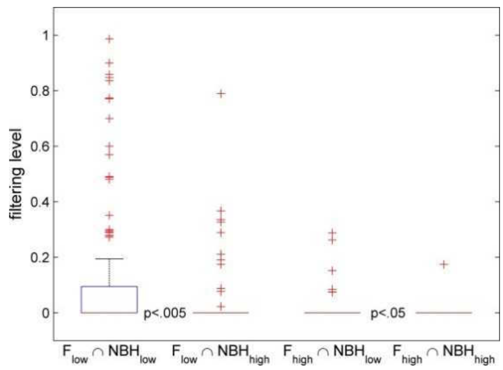

Figure 7: Filtering levels on German word memory traces for serial recall, for four classes of word-frequency by neighbourhood-entropy bins.

However, neighbours are not always helpful. If a word to be recalled is associated with high-frequency neighbours, these neighbours tend to strongly interfere with recall, eventually leading the map astray. The lower the frequency of the target word is, the more prone to interference from competing neighbours it will be, as shown in Fig. 7 for German verbs, where low-frequency words in low-entropy neighbourhoods ($F_{low} \cap NBH_{low}$) appear to require a significantly higher level of filtering than words in high-entropy neighbourhoods do.

3 Concluding remarks

Wordlikeness is a fundamental determinant of lexical organisation and access. Two quantitative measures of wordlikeness, namely *n*-gram probability and neighbourhood density, relate to important dimensions of lexical organisation: the syntagmatic (or horizontal) dimension, which controls the level of predictability and entrenchment of a serial memory trace, and the paradigmatic (or vertical) dimension, which controls the number of neighbours that are co-activated by the target word. The two dimensions are nicely captured by TSOMs, allowing the investigation of their dynamic interaction.

In accessing and recalling a target word, a large pool of neighbours can be an advantage, since they tend to support patterns of activation that are shared by the target word. However, their help may turn out to interfere with recall, if the connection strength of one or more neighbours is overwhelmingly higher than that of the target. Deeply entrenched friends eventually become competitors. This dynamic establishes a nice connection with paradigm acquisition, where a uniform distribution of paradigm members is helpful in spreading morphotactic information and speed up acquisition, and paradigmatically-related forms in skewed distributions compete with one another (Marzi et al. 2014). We argue that both neighbourhood and morphological effects are the result of predictive (syntagmatic) activation and competitive (paradigmatic) co-activation of parallel processing nodes in densely interconnected networks.

As a final qualification, our experiments illustrate the dynamic of activation and storage of letter strings, with no information about morphological content. They provide evidence of the first access stages of early lexical processing, where strategies of automatic segmentation are sensitive to possibly apparent morphological information (Post et al. 2008). Nonetheless, our data suggest that perception of wordlikeness and morphological structure can be accounted for by a common pool of principles governing the organisation of long-term memories for time series.

References

Baddeley, A. D. (1964). Immediate memory and the "perception" of letter sequences. *Quarterly Journal of Experimental Psychology*, 16, 364–367.

Gathercole, S. E., C. Willis, H. Emslie & A.D. Baddeley (1991). The influence of syllables and wordlikeness on children's repetition of nonwords. *Applied Psycholinguistics*, 12, 349–367.

Bard, E.G. (1990). Competition, lateral inhibition, and frequency: Comments on the chapters of Frauenfelder and Peeters, Marslen-Wilson, and others. In G. T. M. Altmann (ed.), *Cognitive models of speech processing: Psycholinguistic and computational perspectives,* 185-210. MIT Press.

Bailey, T. M., & U. Hahn (2001). Determinants of wordlikeness: Phonotactics or lexical neighborhoods? *Journal of Memory and Language*, 44(4), 568-591.

Baayen, R., L. Feldman & R. Schreuder (2006). Morphological influences on the recognition of monosyllabic monomorphemic words. *Journal of Memory and Language*, 53, 496–512.

Coltheart, M., K. Rastle, C. Perry, R. Langdon & J. Ziegler (2001). DRC: A Dual Route Cascaded model of visual word recognition and reading aloud. *Psychological Review*, 108, 204-256.

Davis, C. J. (2010). The spatial coding model of visual word identification. *Psychological Review*, 117.3, pp. 713-758.

Ferro, M., D. Ognibene, G. Pezzulo & V. Pirrelli (2010). Reading as active sensing: a computational model of gaze planning in word recognition. *Frontiers in Neurorobotics*, 4(6), 1-16.

Goldinger, S. D., P. Luce, & D. Pisoni (1989). Priming lexical neighbours of spoken words: Effects of competition and inhibition. *Journal of Memory & Language,* 28, 501-518.

Grainger, J. & W. van Heuven (2003). Modeling letter position coding in printed word perception. *The mental lexicon*, 1-24. New York, Nova Science.

Harm, M.W. & M.S. Seidenberg (1999). Phonology, Reading Acquisition and Dyslexia: Insights from Connectionist Models. *Psychological Review,* 106(3), 491-528.

Henson, R.N. (1998). Short-term memory for serial order: The start-end model. *Cognitive Psychology*, 36, 73-137.

Huntsman, L.A. & S.D. Lima (2002). Orthographic Neighbors and Visual Word Recognition. *Journal of Psycholinguistic Research*, 31, 289-306.

Kohonen, T. (2001). Self-Organizing Maps. Heidelberg, Springer-Verlag.

Koutnik, J. (2007). Inductive Modelling of Temporal Sequences by Means of Self-organization. In Proceeding of International Workshop on Inductive Modelling (IWIM 2007), Prague, 269-277.

Luce, P. & D. Pisoni (1998). Recognizing Spoken Words: The Neighborhood Activation Model. *Ear and hearing*, 19(1), 1-36.

Luce, P., D. Pisoni & S. D.Goldinger (1990). Similarity neighborhoods of spoken words. In G. T. M. Altmann (Ed.), *Cognitive models of speech processing: Psycholinguistic and computational perspectives,* 122-147. Cambridge, MA: MIT Press.

Marzi C., M. Ferro & V. Pirrelli. (2012a). Prediction and Generalisation in Word Processing and Storage. In 8thMediterranean Morphology Meeting Proceedings on Morphology and the architecture of the grammar, 113-130.

Marzi C., M. Ferro & V. Pirrelli. (2012b). Word alignment and paradigm induction. *Lingue e Linguaggio* XI, 2. 251-274. Bologna: Il Mulino.

Marzi C., M. Ferro & V. Pirrelli. (2014). Morphological structure through lexical parsability. *Lingue e Linguaggio*, 13(2), forthcoming.

McClelland, J.L. & D.E. Rumelhart (1981). An interactive activation model of context effects in letter perception: Part 1. An account of Basic Findings. *Psychological Review*, 88, 375-407.

Montemagni, S. *et al.* (2003). Building the Italian syntactic-semantic treebank. In Abeillè, A. (ed.) *Building and Using Parsed Corpora*, 189–210. Dordrecht Kluwer.

Page, M.P.A. & D. Norris (1998). The primacy model: a new model of immediate serial recall. *Psychological Review*, 105, pp. 761-781.

Perry, C., J.C. Ziegler & M. Zorzi (2007). Nested incremental modeling in the development of computational theories: The CDP+ model of reading aloud. *Psychological Review*, 114(2), 273-315.

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

Plaut, D.C., J.L. McClelland, M.S. Seidenberg & K. Patterson (1996). Understanding normal and impaired word reading: Computational principles in quasi-regular domains. *Psychological Review,* 103, 56-115.

Post, B., W. Marslen-Wilson, W., B. Randall & L.K. Tyler (2008). The processing of English regular inflections: Phonological cues to morphological structure. *Cognition* 109 1-17.