A Self-Organizing Model of the Bilingual Reading System

Stéphan Tulkens, Dominiek Sandra, Walter Daelemans
name.surname@uantwerpen.be

CLiPS
Computational Linguistics & Psycholinguistics
University of Antwerp

Introduction

We present **Global Space**, a model of the mental lexicon, based on Self-Organizing Maps (SOM) [1].
The model is trained on word representations which consist of the concatenation of a corresponding **orthographic** and **phonological** vector.
As such, the model does not assume independence between, or the existence of, separate phonological and orthographic processing modules.
The model is trained on these concatenated representations, but only receives the orthographic part of the vector during regular functioning.
The current work focuses on establishing the basic function of the model and its theoretical underpinnings.

Dynamic Systems Theory

One of the assumptions of Dynamic Systems Theory (DST) is that cognition can be modeled as a trajectory through space, where the path of the trajectory is dependent on the current input and the recent history[2].
This is an attractive framework for models of word reading, because it provides us with a way of accounting for both priming and context effects.

Time Dependence

A standard SOM does not have time dependence, and hence does not provide us with a good model of the time course of word reading.
We add time dependence to the SOM by defining a state vector s^n, where n is the number of nodes on the SOM. s defines the activation of each node at the current time t.

s is updated for each t, as follows: for a given input x, we calculate the map response as the euclidean distance between the weight matrix W and input x. This response is fed through an activation function:

$$a = exp(-||W - x||_2)$$
$$a = a - (mean(a) + std(a))$$

At the same time, using s, we construct an exemplar, r, from the map activations by multiplying the state vector by the weight matrix, and taking the column-wise mean of the resulting matrix.

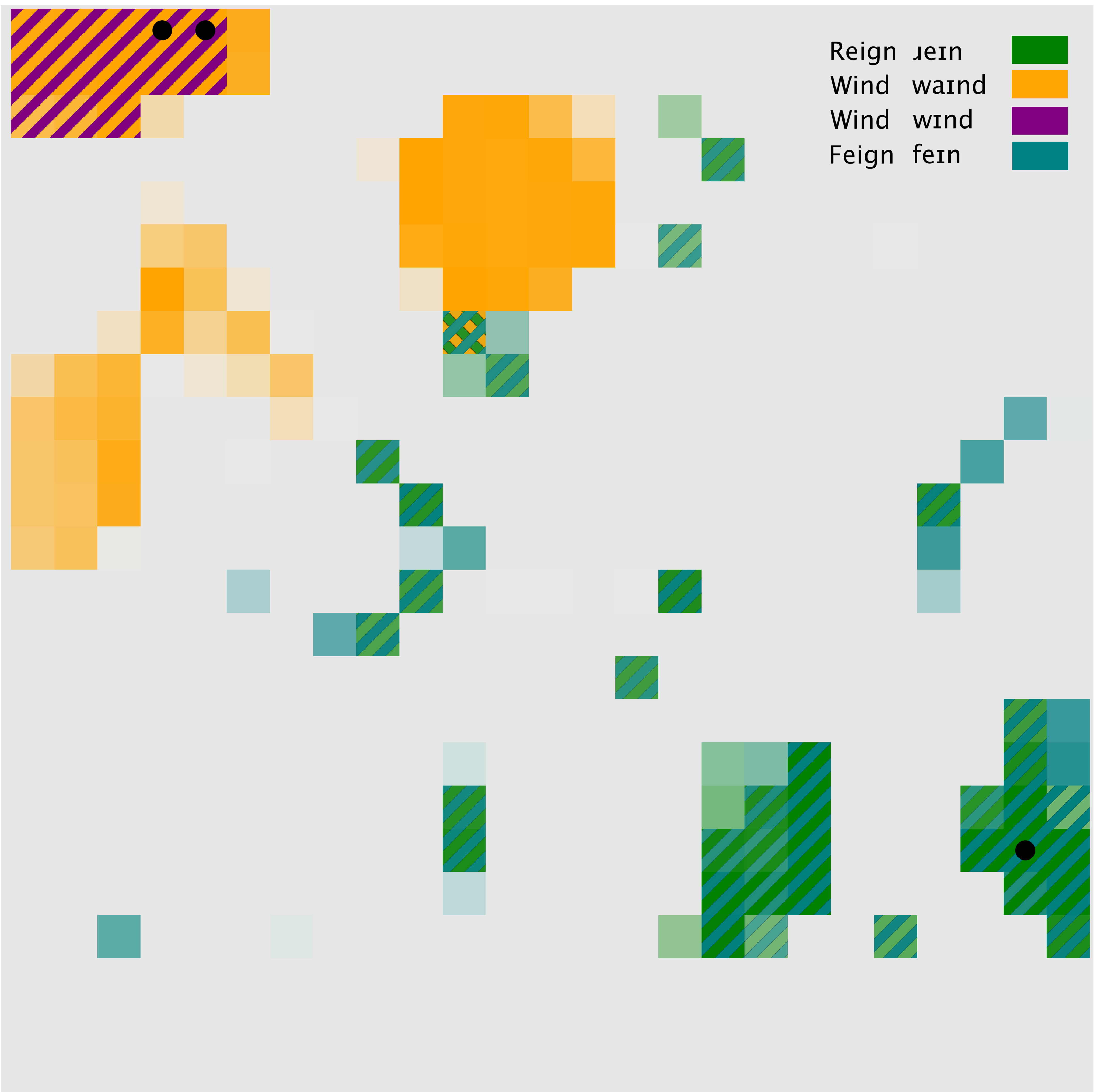
$$r_j = \frac{1}{N}\sum_{n=0}^{N} s_n W_{jn}$$

r is then fed through the same activation function, and added to a, resulting in a vector d^n. Δs is then computed as follows:

$$\Delta s_n = \begin{cases} (d_n(1 - (\frac{s_n^t}{2}))) & \text{if } d_n > 0 \\ (d_n S_n^t) & \text{if } d_n \leq 0 \end{cases}$$

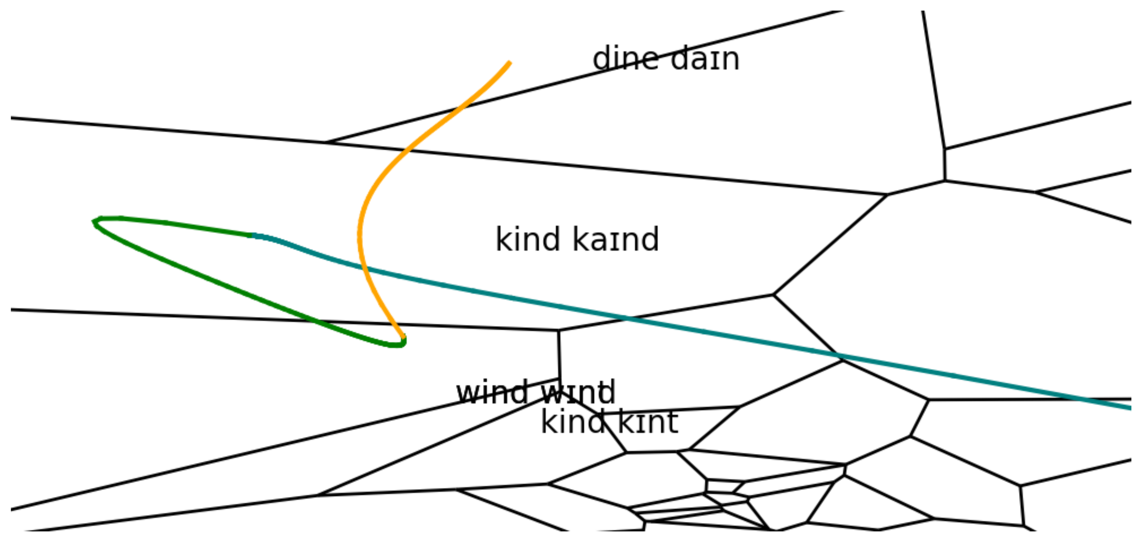
This update rule ensures that s is bounded to [0, 2], and that updates get smaller as s approaches the upper or lower bound. This procedure introduces time dependence because updates gradually slows down and converge to a stable state. The rate of convergence is dependent on various factors, including frequency and number of neighbors.

Distributed representations



We define the **distributed representation** for each word as s after this word converges using the time dependence procedure described above.
This overcomes a key shortcoming of a SOM model of word reading: **storage**. Very similar words often get assigned the same neuron, which, by most accounts, makes these indistinguishable by the model.
Figure 1 shows that representations of words with the same center still have dissimilar attractors.

Phase Space



The distributed representations allow us to define a **phase space** over all the words in the mental lexicon.
Each colored line denotes the trajectory of the state of the model given only some orthographic input.

Aside from theoretical considerations, using a distributed representation also improves accuracy.

	Accuracy
standard SOM	.725
Distributed	.815

Materials

The model was trained on a set of 108 Dutch and English words, selected from `CELEX`
SOM toolkit:
`github.com/stephantul/somber`
System and experiments:
`github.com/stephantul/wavesom`

References

[1] Teuvo Kohonen. The self-organizing map. *Neurocomputing*, 21(1):1–6, 1998.

[2] Robert F Port and Timothy Van Gelder. *Mind as motion: Explorations in the dynamics of cognition*. MIT press, 1995.