Doursat, R. & Petitot, J. (2005a) Bridging the gap between vision and language: A morphodynamical model of spatial categories.
Proceedings of the International Joint Conference on Neural Networks (IJCNN 2005), July 31-August 4, 2005, Montréal, Canada.

Bridging the Gap Between Vision and Language: A Morphodynamical Model of Spatial Categories

René Doursat
Goodman Brain Computation Lab
University of Nevada, Reno, NV 89557, USA
E-mail: doursat@unr.edu
Web: www.cse.unr.edu/~doursat

Jean Petitot
CREA, Ecole Polytechnique
1, rue Descartes, 75005 Paris, France
E-mail: petitot@poly.polytechnique.fr
Web: www.crea.polytechnique.fr

Abstract— **We propose a spiking neural network for mapping the infinity of schematic visual scenes to a small set of semantic symbols. According to cognitive linguistics, spatial prepositions such as 'in' or 'above' are neutral toward the shape and size of objects. We suggest that they correspond to *morphodynamical transforms* erasing details and creating virtual structures (boundaries, skeleton). These *singularities* arise from a large-scale lattice of coupled excitable units exhibiting pattern formation through spatiotemporal order, especially *traveling waves*. Our model addresses the fundamental cognitive mechanisms of spatial *schematization* and *categorization*, which mediate vision and language and are crucial in designing intelligent systems.**

I. Introduction

A. Spatial Categorization

How can the same English relationship 'in' apply to scenes as different as "the shoe in the box" (small, hollow, closed volume), "the bird in the tree" (large, dense, open volume) or "the fruit in the bowl" (curved surface)? What is the common 'across' invariant behind "he swam across the lake" (smooth trajectory, irregular surface) and "the fly zigzagged across the hall" (broken trajectory, regular volume)? How can language, especially its spatial elements, be so insensitive to wide topological and morphological differences among visual percepts? In short, how does language drastically simplify information and *categorize*?

This study examines the link between the spatial structure of visual scenes and their linguistic descriptions. We propose a computational and spiking neural model aimed at mapping the endless diversity of schematic visual scenes to a small set of standard semantic labels. Contemporary theories of metaphor[1] have demonstrated that terms with spatiotemporal content are highly polysemous. We therefore restrict our scope to relatively "homogeneous" subcategories, or *protosemantic* concepts. For example[2], "the cat in_1 the house" and "the bird in_2 the tree" should be treated separately from "the flower in_3 the vase" or "the crack in_4 the vase". Yet, even within a monosemantic subconcept, there remains the great difficulty of relating the infinite continuum of shape diversity to a discrete symbol. This task addresses the brain's fundamental mechanisms of schematization and categorization, which mediate vision and language and play a critical role in the design of intelligent systems. In recent years, neural networks and reactive robotics have successfully challenged symbolic artificial intelligence and formal logic by showing how complex behavior, object recognition or learning can arise from simple dynamic control loops. Yet, for further progress in this direction there is a need for greater generalization capabilities at more abstract levels of spatial representation. The different trajectories followed by a robot could be categorized under a few invariant tasks; the different visual scenes captured by its sensors, under a few prototypical situations. Categorization-enabled agents could then interact with other agents or humans at this emergent symbolic level, i.e., the level typical of a natural language.

B. Cognitive Linguistics

Through another recent shift of paradigm, the formalist view that language is functionally autonomous was revised by a set of works[1,3,4,5] collectively named *cognitive linguistics*, for which language is much rather "embodied" in perception, action and inner conceptual representations. In this stance, generative grammar models have been superseded by studies of the *interdependence of language and perceptual reality* and how these two systems influence each other's organization. For example, in "I am in/on/far from the street", the elements 'in', 'on' and 'far from' contrastively construe the street as a volume, a surface or a reference point. Formal syntax models, for their part, make the street a mere symbol, irrespective of its spatial features. These alternative linguistic orientations have naturally converged with major advances in perception and image analysis, both in neurophysiology and machine vision.

Cognitive linguistics is directly preoccupied with meaning and categorization. Refuting the distinction between syntax and semantics, it postulates a *conceptual level* of representation, where language, perception and action become compatible[5]. It also remarkably revived the Gestalt approach calling into question the traditional roles assigned to perception—as a faculty only dealing with object shapes, and language—as a faculty only dealing with relations between objects. Whereas in logical theories of language "things" are already individuated symbols and "relations" are abstract links connecting these symbols, in the Gestaltist or *mereological* conception things and relations constitute wholes: relations are not given for granted but emerge together with the objects through segmentation and transformation. In fact, for Gestalt linguistics the perceptual background is a true figure, actively structured by the *grammatical* elements in a spatial and morphological way that is radically different from symbolic relations.

C. Linguistic Topology

Reinforcing this view, there is extensive evidence that grammatical elements select certain morphological features from the perceptual data and ignore others[3]. These elements are largely invariant with respect to the dimensions and detailed shapes of objects and trajectories. For example, "the caterpillar crawled up along the filament/flagpole/ redwood tree"[3] shows that the English preposition 'along' construes the background object only as its central axis and is indifferent to its girth. Similarly, the 'in' and 'across' examples cited in section I.A indicate neutrality toward the topological diversity of the container. On the other hand, the domain of applicability of grammatical elements can also be sensitive to metric ratios: "he swam across the pool lane" implies a swimmer's trajectory crossing the rectangle of water parallel to its short side, not its long side.

This core invariance of spatial meaning is sometimes referred to as the "linguistic form of topology"[3] or "cognitive topology"[1]. Yet, language can also display a greater power of generalization than mathematical topology (example 'in') while in other cases it can preserve metric aspects and constrain distortions much more strictly than topology (example 'across'). This apparent paradox constitutes one of the great puzzles of spatial cognition.

II. Morphodynamical Cellular Automata

A. Perceptual-Semantic Machine

One step toward solving this puzzle is to propose that spatial prepositions like 'in', 'above', 'across', etc., fundamentally amount to *morphodynamical transforms*, i.e., transforms creating a morphology that evolves temporally. The above findings could be explained by transformation routines performing a drastic, yet targeted simplification of the geometric data. These routines essentially (a) erase details and (b) create new, virtual structures or *singularities* (e.g., influence zones, fictive boundaries, skeletons, intersection points, force fields, etc.) that were not originally part of the scene but are ultimately revealing of the conveyed meaning. At the core of such transformations are (1) dynamical, object-centered expansion processes, akin to diffusion or propagation, and (2) routines detecting the singularities created by these processes. A key idea is that singularities encode a lot of the image's geometrical information in an extremely compact and localized manner.

In a first attempt[6] to test this hypothesis, we have designed a morphodynamical image-processing software or "perceptual-semantic" machine. In input we present schematic scenes, assumed to be the outcome of the early stages of visual processing. The actors of the semantic relationship, the figure and ground, or *trajector* (*TR*) and *landmark* (*LM*)[4], are presegmented. In output we obtain symbolic features containing global information about the respective spatial positions of the actors. These features are *protosemantic*, i.e., low-level compared to linguistic-cultural categories but high-level compared to local visual features. They correspond to subcategorical "islands" in a ramified, prototype-based category. An additional classification module based on learning could be used to implement the final step from the protofeatures to the full-fledged prepositions, however we will not address learning issues here. In the remainder of this article we focus on the core of the system, the "morphodynamical engine" that transforms images into protosemantic symbols and creates the bridge from vision to language.

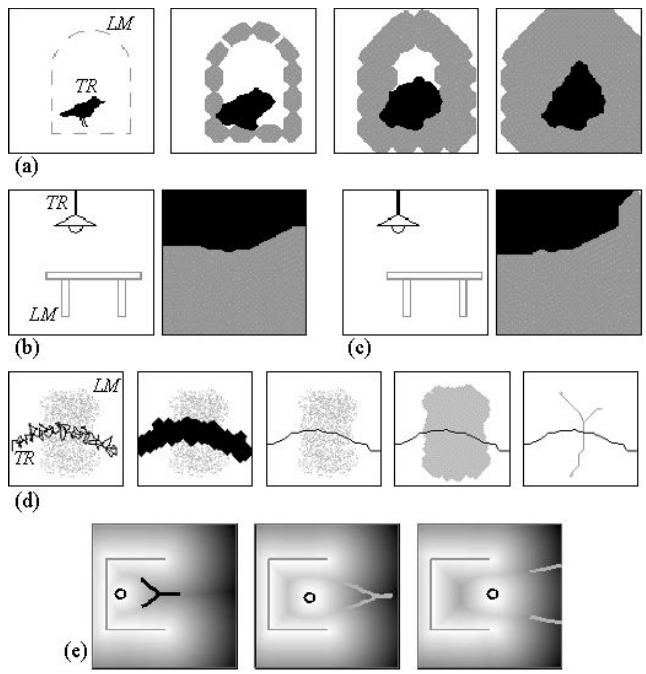

Fig. 1. Morphodynamical Detection of Protosemantic Schemas. (a) 'In': the bird's expansion is blocked by the cage from reaching the borders, despite the holes. (b) 'Above': the lamp cannot reach the bottom because of the table, (c) even when not aligned vertically; (d) 'Across': the zig-zagged path and textured domain are skeleton-transverse. (e) 'Out of': the singular point on the ball's and box's influence boundaries disappears.

B. Examples of Morphodynamical Routines

1) Containment of Influence Zones in Proto-'in' and -'above': For example, using 2-D cellular automata (CA) with a simple nearest-neighbor diffusion rule, one way to define *TR* 'in' *LM* ("the bird in the cage") could be to detect whether an isotropic expansion of *TR* is contained within *LM*'s own expansion, i.e., check that no *TR*-induced activity reaches the image borders (Fig. 1a). Similarly, *TR* 'above' *LM* ("the lamp above the table") could correspond to *TR*'s expansion stopped by *LM* from reaching the bottom (Fig. 1b-c: the scene has morphed into a roughly horizontal boundary line, revealing an 'above' relation).

2) Transversality of Skeleton in Proto-'across': For *TR* 'across' *LM* (Fig. 1d), *TR*'s and *LM*'s medial axes should be transverse, i.e., intersect nontangentially. These axes are obtained by *skeletonization*, an inward expansion eroding a shape to its axis, possibly preceded by *tubification*, a limited outward expansion erasing irrelevant texture details.

3) Dynamic Bifurcation in Proto-'out of': Fig. 1e illustrates another scenario, *TR* 'out' of *LM* ("the ball out of the box"). In this case, the singular point on the influence boundaries of *TR* and *LM* (triple intersection highlighted by the thick lines) accelerates away from *TR* and eventually disappears as *TR* exits the interior of *LM*. This is a very robust bifurcation phenomenon, too, independent of the detailed shapes of the objects or *TR*'s trajectory.

C. Expansion-Based Topology

We suggest that the brain might rely on dynamical patterns of activity of this kind to perform invariant spatial categorization. There is experimental evidence that the visual system effectively constructs the symmetry axis of shapes[7,8]. We therefore postulate the following principles: (i) objects have a tendency to occupy the whole space; (ii) objects are obstacles to each other's expansion. Through the action of structuring routines the common space shared by the objects is divided into *influence zones*. Image elements cooperate to propagate activity across the field and inhibit activity from other sources. This creates singularities, such as boundaries and intersection points, which constitute the characteristic "signature" of the spatial relationship[9]. The transformation routines thus considerably reduce the dimensionality of the input space, literally "boiling down" the input images to a few key features.

III. Waves In Spiking Neural Networks

A. Dynamic Pattern Formation in Excitable Media

Elaborating upon this first morphodynamical model, we now establish a parallel with neural modeling. Our main hypothesis is that the transition from analog to symbolic representations of space might be neurally implemented by *traveling waves in a large-scale network of coupled spiking units*, via the expansion processes discussed above (see Fig. 2). There is a vast cross-disciplinary literature, revived in the 1970's[10], on the emergence of ordered patterns in *excitable media* and coupled oscillators. Traveling or kinematic waves are also a frequent phenomenon in nonlinear chemical reactions or multicellular structures[11], such as slime mold aggregation, heart tissue activity, or embryonic pattern formation. Across various dynamics and architectures, these systems have in common the ability to reach a critical state from which they can rapidly bifurcate between randomness or chaos and ordered activity. To this extent they can be compared to "sensitive plates", as certain external patterns of initial conditions (chemical concentrations, food, electrical stimuli) can quickly trigger internal patterns of collective response from the units.

We explore the same idea in the case of an input image impinging on a layer of neurons and draw a link between the produced response and categorical invariance. In the framework proposed here, a visual input is classified by the qualitative behavior of the system, i.e., the presence or absence of certain singularities in the response patterns.

B. Spatiotemporal Patterns in Neural Networks

During the past two decades, a growing number of neurophysiological recordings have revealed precise and reproducible temporal correlations among neural signals and related them with behavior[12,13,14,15]. *Temporal coding*[16] is now recognized as a major mode of neural transmission, together with average rate coding. In particular, quick onsets of transitory phase locking have been shown to play a role in the communication among cortical neurons engaged in a perceptual task[13].

While most experiments and models involving neural synchronization were based on zero-phase locking among coupled oscillators[17,18], *delayed* correlations have also been observed[12], suggesting *nonzero*-phase locking modes of organization that correspond to reproducible rhythms or waves. In another proposal[19], waves are generated by *synfire chains*[12,20] and construed as the physical basis for elementary "building blocks" composing more complex cognitive objects. These patterns exhibit *compositional* properties, as two waves simultaneously propagating on two chains can lock and merge into one single wave by growing cross-connections between the chains (in a "zipper" fashion). According to this theory, spatiotemporal patterns would then be analogous to folded proteins that bind through conformational interactions.

In the present work, we propose the emergence of wave patterns on regular 2-D lattices of coupled oscillators, which implement the expansion dynamics of the morphodynamical spatial transformations. Therefore, compared to the traditional blocks of synchronization, i.e., phase plateaus often used in segmentation models, we are interested in traveling waves, i.e., phase *gradients*.

C. Wave Propagation and SKIZ

A possible neural implementation of the morphodynamical engine at the core of our model relies on a network of individual spiking neurons, or local groups of spiking neurons (e.g., columns), arranged in a 2-D lattice similar to topographic cortical areas, with nearest-neighbor coupling. Each unit obeys a system of differential equations that exhibit regular oscillatory behavior in a certain region of parameter space. Various combinations of oscillatory dynamics (relaxation, stochastic, reaction-diffusion, pulse-coupled, etc.) and parameters (frequency distribution, coupling strength, etc.) are able to produce waveform activity, however it is beyond the scope of the present work to discuss their respective merits. We want here to point out the generality of the wave propagation phenomenon, rather than its dependence on a specific model. For practical purposes, we use Bonhoeffer-van der Pol (BvP) relaxation oscillators[21] locally coupled to each other within a small radius. Parameters are tuned so that individual units are close to a bifurcation in phase space between a fixed point and a limit cycle, i.e., one spike emission. They are *excitable* in the sense that a small stimulus causes them to jump out of the fixed point and orbit the limit cycle, during which they cannot be disturbed again.

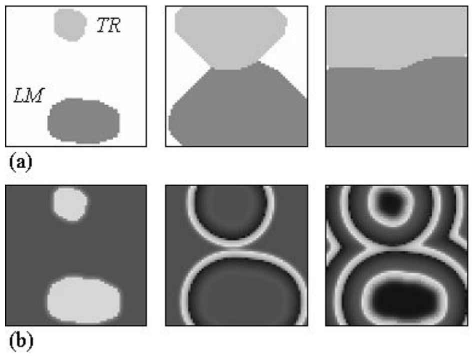

Fig. 2. Realizing Morphodynamical Routines in a Spiking Neural Network. (a) SKIZ obtained by diffusion in a 64x64 CA, as in Fig. 1. (b) Same SKIZ obtained by traveling waves on a 64x64 lattice of coupled BvP oscillators with connectivity radius 2.3 (potential shown in gray levels). An initial input image generates traveling waves in the network.

Fig. 2b shows waves of excitation in a network of coupled BvP units created by the schematic scene "a small blob above a large blob". Starting with uniform resting potentials or a weak level of stochastic firing, the initial impulse triggered by the input image creates fronts of activity that propagate away from the object contours and collide at the boundary between the objects. These fronts are "grass-fire" traveling waves, i.e., single-spike bands followed by refraction and reproducing only as long as the input is applied. Nonlinear waves of this type annihilate when they meet, instead of adding up. Fig. 2a shows the same influence zones obtained by mutual expansion, as seen in the CA model (Fig. 1b-c). In both cases, the border line is the *skeleton by influence zones* or *SKIZ*. It also corresponds to what is usually called the *medial axis*[22] or *shock graph*[23], applied here to the objects' complementary set. As noted previously, there is convincing perceptual and neural evidence for the significant role played by the medial axis and propagation in vision[7].

The dynamics of coupled spiking units (Fig. 2b) is richer than the morphodynamical model (Fig. 1, Fig. 2a) because it contains specific patterns of activity that are absent from a static geometric line. In particular, the wave fronts highlight a secondary flow of propagation *along* the SKIZ line, which travels away from the focal shock point with decreasing speed on either branch[22,7]. The focal point (where the bright band is at its thinnest in Fig. 2b) is the closest point to both objects and constitutes a *local optimum* along the SKIZ. While a great variety of object pairs produce the same static SKIZ, the speed and direction of flow along the SKIZ vary with the objects' relative curvature and proximity to each other. For example, a vertical SKIZ segment flows outward between brackets facing their convex sides)|(, whereas it flows inward between reversed brackets (|). This information is revealed by wave propagation and can be exploited for refined classification schemes.

D. SKIZ Signature Detection

In order to detect the focal points and flow characteristics (speed, direction) of the SKIZ, we propose in this model additional layers of neural cells similar to the so-called "complex cells" of the visual system. These detector cells receive afferent contacts from local neighborhoods in the input layer and respond to segments of moving wave bands, with selectivity to their orientation and direction. More precisely, Fig. 3 shows the detection of a protosemantic 'above' situation. The spiking neural network presented here is a three-tier architecture comprising: (a) two input layers, (b) two middle layers of orientation and direction-selective "*D*" cells, and (c) four top layers of coincidence "*C*" cells responding to specific pairwise combinations of *D* cells. Note that these are not literally cortical layers but might rather correspond to functionally distinct cortical areas, or subnetworks thereof.

In this particular setup, the original input layer is split into two layers, L_{TR} and L_{LM} (Fig. 3a), each holding one presegmented component of the scene. A number of models have shown that segmentation from contiguity can arise on a lattice of coupled oscillators through temporal tagging (zero-phase synchronization). We take these results as the starting point of our simulations and assume that the initial segments are forwarded to two separate sublayers, where they independently generate a single wave front (Fig. 3a is a thresholded version of Fig. 2b). Here, the waves created by *TR* and *LM* do not actually collide. Rather, the region where they coincide "vertically" (viewing layers L_{TR} and L_{LM} superimposed) can be captured by higher feature cells in two direction-selective layers, D_{TR} and D_{LM} (Fig. 3b), and four coincidence-detection layers, $C_{1\ldots4}$ (Fig. 3c). D_{TR} receives afferents only from L_{TR}, and D_{LM} only from L_{LM}. Layers C_i are connected to both D_{TR} and D_{LM} through split receptive fields: half of the afferent connections of a *C* cell originate from a half-disc in D_{TR} and the other half from its complementary half-disc in D_{LM} (details below).

In the intermediate *D* layers, each point contains a family of cells, or "jet", similar to multiscale Gabor filters. Viewing the traveling waves in layers *L* as moving bars of activity, each *D* cell is sensitive to a combination of bar width λ, speed v, orientation θ and direction of movement φ. In the simple wave dynamics of the *L* layers, λ and v are approximately uniform. Therefore, a jet of *D* cells is in fact single-scale and indexed by one parameter $\theta = 0\ldots2\pi$, with the convention that $\varphi = \theta + \pi/2$. Typically, 8 cells with orientations $\theta = k\pi/4$ are sufficient. For ease of view, both *D* layers in Fig. 3b are displayed as 8 sublayers (small squares) with $k = 8, 7, \ldots, 1$ starting north, going clockwise. Each sublayer $D_{TR}(\theta)$ thus detects a portion of the traveling wave in L_{TR} (same with *LM*). Realistic neurobiological architectures generally implement direction-selectivity using inhibitory cells and transmission delays.

In our simplified model, a $D(\theta)$ cell is a "cylindrical" filter, i.e., a temporal stack of discs containing a moving bar at angle θ (see illustration at center of D layers): it sums potentials from afferent L-layer spikes spatially and temporally, and fires itself a spike above a certain threshold.

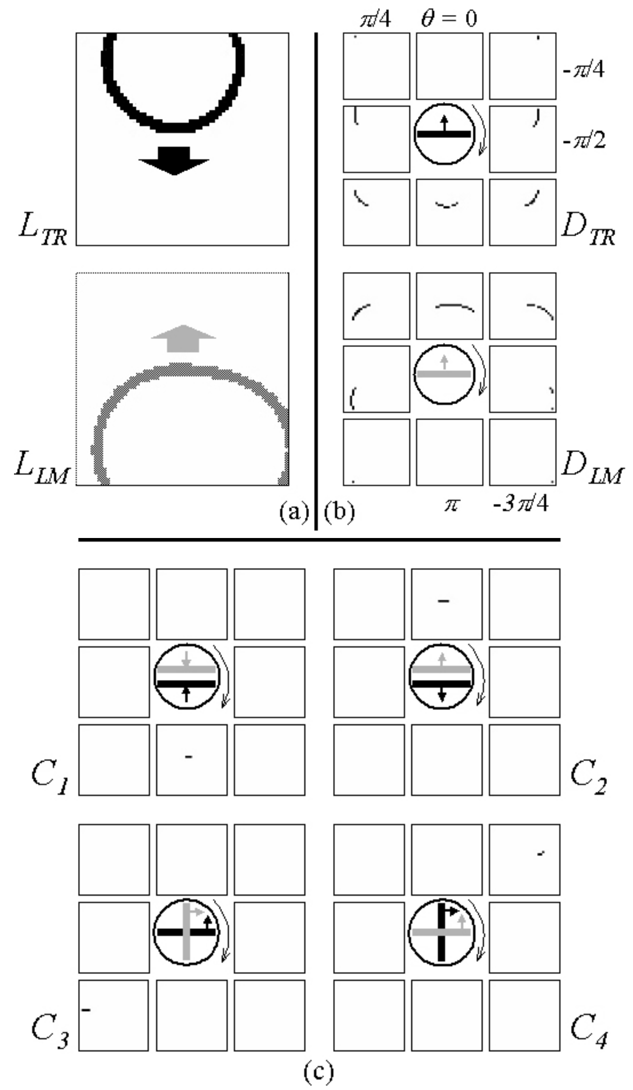

Fig. 3. SKIZ Detection in a Three-Tier Spiking Neural Network Architecture. (a) Input layers. (b) Orientation and direction-selective cells. (c) Pairwise coincidence cells (see text).

Among the four top layers (Fig. 3c), C_1 detects converging parallel wave fronts, C_2 detects diverging parallel wave fronts, and C_3 and C_4 detect crossing perpendicular wave fronts. Like the D layers, each C_i layer is subdivided into 8 orientation sublayers $C_i(\theta)$. Each cell in $C_1(\theta)$ is connected to a half-disc neighborhood in $D_{TR}(\theta)$ and the complementary half-disc in $D_{LM}(\theta - \pi)$, where the half-disc separation is at angle θ. For example: a cell in $C_1(0)$ (illustrated in the center icon of C_1) receives afferents from a horizontal bar of $D_{TR}(0)$ cells in the lower half of its receptive field, and a bar of $D_{LM}(\pi)$ cells in the upper half (for $C_2(0)$, swap TR/LM and upper/lower). Similarly, $C_3(\theta)$ cells are half-connected to $D_{TR}(\theta)$ and $D_{LM}(\theta - \pi/2)$, with orthogonal bars (swapping TR/LM for C_4). The net output of this hierarchical arrangement is a *signature* of coincidence detection features providing a *very sparse coding* of the original spatial scene. The input scene "a blob above a blob" is eventually reduced to a handful of active cells in a single orientation sublayer $C_i(\theta)$ for each C_i: $C_1(\pi)$, $C_2(0)$, $C_3(3\pi/4)$ and $C_4(-\pi/4)$ (Fig. 3c). Note that the activity traces shown here are persistent: in reality, C_1 cells fire first when the two wave fronts meet, then C_2 cells when the fronts separate again, and finally C_3 and C_4 cells in close succession when the arms of the waves cross.

In summary, the active cells in C_1 and C_2 reveal the focal point of the SKIZ, which is the primary information about the scene, while C_3 and C_4 reveal the outward flow on the SKIZ branches, which can be used to distinguish among similar but nonequivalent concepts. This sparse SKIZ signature is at the same time characteristic of the spatial relationship and largely insensitive to shape details. For example: 'below' yields $C_1(0)$ and $C_2(\pi)$, while 'on-top-of' produces the same as 'above' but without C_1 activity because TR and LM are contiguous (wave fronts can only separate at a contact point, not join).

IV. Discussion

We have proposed a dynamical approach to cognitive linguistics drawing from morphological CA and spiking neural networks. We suggest that spatial semantic categorization can be supported by expansion-based dynamics, such as activity diffusion or wave propagation. Admittedly, the few results we have presented here are not particularly surprising given the relative simplicity and artificiality of the models. However, we hope that this preliminary study will be a starting point for more efficient or plausible network architectures exploring the interface between high-level vision and symbolic knowledge.

A. Original Points of this Proposal

1) Bringing Large-Scale Dynamical Systems to Cognitive Linguistics: Despite their deep insights into the conceptual and spatial organization of language, cognitive grammars still lack mathematical and computational foundations. Our project is among few attempts to import spatiotemporal connectionist models into linguistics and spatial cognition. Other authors[24,25] pursuing the same goal use small "hybrid" artificial neural networks, where nodes already carry geometrical or symbolic features. We work at the fine-grained level of numerous spatially arranged units.

2) Addressing Semantics in CA and Neural Nets: Conversely, our work is also an original proposal to apply large-scale lattices of CA or neurons to high-level semantic feature extraction. These bottom-up systems are usually exploited for low-level image processing or visual cortical modeling, or both, e.g., Pulse-Coupled Neural Networks[26] or Cellular Neural Networks[27]. Shock graphs and medial axes are also advocated in computer vision models of object recognition[23,28,29], but with the concern to preserve and match object shapes, not erase them. Collision-based wave dynamics was also proposed for logic-gate computing[30].

3) Promoting Pattern Formation in Neural Modeling: Self-organized, emergent processes of pattern formation, or *morphogenesis*, are ubiquitous in nature (stripes, spots, branches, etc.). As a complex system, the brain produces "cognitive forms", too, but instead of spatial arrangements of molecules or cells, these forms are made of *spatiotemporal patterns* of neural activities (synchronization, correlations, waves, etc.). In contrast to other biological domains, however, pattern formation in large-scale neural networks has attracted only few authors[31]. This is probably because precise rhythms involving a large number of neurons are still experimentally difficult to detect, hence not yet proven to play a central role.

4) Suggesting Wave Dynamics in Neural Organization: Indeed, fast waveform activity on the 1-ms timescale in random[19] or regular[32] networks has been much less explored than synchronization, when addressing segmentation or the "binding problem". Along with other authors[19], we contend that waves open a richer space of temporal coding suitable to general *mesoscopic* neural modeling, i.e., the intermediate level of organization between microscopic neural activities and macroscopic representations. At one end (AI), high-level formal models manipulate symbols and composition rules but do not address their fine-grain internal structure. At the other end (neural nets), low-level dynamical models study the self-organization of neural activity but their emergent objects (attractor cell assemblies or blobs) still lack structural complexity and compositionality[33]. Waves and other complex spatiotemporal patterns could provide the missing link bridging the gap between these two levels. Furthermore, our hypothesis is that this mesoscopic level corresponds to the central conceptual level postulated by cognitive linguistics.

B. Future Work

Among the multiple directions that this preliminary work is hinting at, we can mention: (i) extending to real-world images; (ii) mapping the protosemantic features to real linguistic elements and, thereby, learning cross-cultural differences (e.g., the English 'on' vs. the German 'auf' and 'an'); (iii) modeling verbal scenarios, in which the singularities created by fast wave activity evolve themselves on a slower timescale (Fig. 1e), drawing from studies about the perception of causality in schematic movies[34].

Acknowledgment

R.D. is grateful to Philip Goodman at UNR for welcoming him in his lab and providing a stimulating environment, and thanks J.P. for his never-failing encouragements.

References

[1] Lakoff, G. (1987). *Women, Fire, and Dangerous Things.* Chicago, IL: University of Chicago Press.
[2] Herskovits, A. (1986). *Language and Spatial Cognition: An Interdisciplinary Study of the Prepositions in English*. Cambridge U. Press.
[3] Talmy, L. (2000). *Toward a Cognitive Semantics. Volume I: Concept Structuring Systems*. Cambridge, MA: The MIT Press.
[4] Langacker, R. (1987). *Foundations of Cognitive Grammar*. Palo Alto, CA: Stanford University Press.
[5] Jackendoff, R. (1983). *Semantics and Cognition*. The MIT Press.
[6] Doursat, R. & Petitot, J. (1997). Modèles dynamiques et linguistique cognitive: vers une sémantique morphologique active. In J. Lorenceau, A. Streri, B. Victorri & Y.-M. Visetti (Eds.), *Actes de la 6è École d'été de l'Association pour la recherche cognitive*. CNRS.
[7] Kimia, B. B. (2003). On the role of medial geometry in human vision. *Journal of Physiology Paris*, *97*(2/3), 155–190.
[8] Lee, T. S. (2003). Computations in the early visual cortex. *Journal of Physiology Paris*, *97*(2/3), 121–139.
[9] Petitot, J. (1995). Morphodynamics and attractor syntax. In T. van Gelder & R. Port (Eds.), *Mind as Motion* (pp. 227–281). MIT Press.
[10] Winfree, A. (1980). *The Geometry of Biological Time*. Springer.
[11] Swinney, H. L. & Krinsky, V. I., Eds. (1991). *Waves and Patterns in Chemical and Biological Media*. Cambridge, MA: The MIT Press.
[12] Abeles, M. (1982). *Local Cortical Circuits: An Electrophysiological Study.* Berlin: Springer-Verlag.
[13] Gray, C. M., König, P., Engel, A. K. & Singer, W. (1989). Oscillatory responses in cat visual cortex exhibit intercolumnar synchronization which reflects global stimulus properties. *Nature*, *338*, 334–337.
[14] O'Keefe, J. & Recce, M.L. (1993). Phase relationship between hippocampal place units and the EEG θ rhythm. *Hippoc.*, *3*, 317–330.
[15] Bialek, W., Rieke, F., de Ruyter van Steveninck, R. & Warland, D. (1991). Reading a neural code. *Science*, *252*, 1854–1857.
[16] von der Malsburg, C. (1981). *The correlation theory of brain function.* Internal Report 81-2. Max Planck Inst. for Biophys. Chemistry.
[17] König, P. & Schillen, T. B. (1991). Stimulus-dependent assembly formation of oscillatory responses. *Neural Computation*, *3*, 155–166.
[18] Campbell, S. & Wang, D. (1996). Synchronization and desynchronization in a network of locally coupled Wilson-Cowan oscillators. *IEEE Transactions on Neural Networks*, *7*, 541–554.
[19] Bienenstock, E. (1995). A model of neocortex. *Network*, *6*, 179–224.
[20] Ikegaya, Y., Aaron, G., Cossart, R., Aronov, D., Lampl, I., Ferster, D. & Yuste, R. (2004). Synfire chains and cortical songs: temporal modules of cortical activity. *Science*, *304*, 559–564.
[21] FitzHugh, R. A. (1961). Impulses and physiological states in theoretical models of nerve membrane. *Biophysical Journal*, *1*, 445–466.
[22] Blum, H. (1973). Biological shape and visual science. *Journal of Theoretical Biology*, *38*, 205–287.
[23] Siddiqi, K., Shokoufandeh, A., Dickinson, S. J. & Zucker, S. W. (1999). Shock graphs and shape matching. *International Journal of Computer Vision*, *35*(1), 13–32.
[24] Regier, T. (1996). *The Human Semantic Potential*. The MIT Press.
[25] Shastri, L. & Ajjanagadde, V. (1993) From simple associations to systematic reasoning. *Behav. and Brain Sciences*, *16*(3), 417–451.
[26] Johnson, J. L. (1994). Pulse-coupled neural nets: translation, rotation, scale, distortion and intensity signal invariance for images. *Applied Optics*, *33*(26), 6239–6253.
[27] Chua, L. & Roska, T. (1998). *Cellular Neural Networks and Visual Computing*. Cambridge, UK: Cambridge University Press.
[28] Zhu, S. C. & Yuille, A. L. (1996). FORMS: A flexible object recognition and modeling system. *Int. J. of Comp. Vision*, *20*(3), 187–212.
[29] Leyton, M. (1992). *Symmetry, Causality, Mind.* The MIT Press.
[30] Adamatzky, A., Ed. (2002). *Collision-based computing.* Springer.
[31] Ermentrout, B. (1998). Neural networks as spatio-temporal pattern-forming systems. *Reports on Progress in Physics*, *61*, 353–430.
[32] Milton, J. G., Chu, P. H. & Cowan, J. D. (1993). Spiral waves in integrate-and-fire neural networks. In *Advances in Neural Information Processing Systems*, *Vol. 5* (pp. 1001–1007). Morgan Kaufmann.
[33] Fodor, J. & Pylyshyn, Z. (1988). Connectionism and cognitive architecture: a critical analysis. *Cognition*, *28*(1/2), 3–71.
[34] Scholl, B. J. & Tremoulet, P. D. (2000). Perceptual causality and animacy. *Trends in Cognitive Science*, *4*(8), 299–309.