Doursat, R. & Petitot, J. (2011) Chapter 3: Relations. In *Cognitive Morphodynamics: Dynamical Morphological Models for Constituency in Perception and Syntax*, by J. Petitot, in collaboration with R. Doursat: pp. 119-170. Peter Lang, ISBN 978-3-0343-0475-7.

# CHAPTER 3

# Relations

## In collaboration with René Doursat

# 1. Introduction

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" (jagged 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*?

The previous chapter addressed the basic concept of *things* and introduced algorithms dealing with image segmentation, perceptual mereology and object constituency. We now turn to the second major component of Langacker's trilogy, the *relations* between things. Relations clearly constitute the most important problem at the core of all theories of language. Ultimately, different theoretical perspectives on syntax will be distinguished on the basis of how they construe relations. It is only after relations are expressed in a mathematical form that *processes* can be modeled as temporal evolutions of relations and *events* as changes occurring during these processes.

## 1.1. The gestaltic conception of relations

The various theoretical paradigms of language are broadly divided between formal or symbolic conceptions of things and relations, on the one hand, and what we here call "gestaltic" conceptions, on the other hand. In the formal framework, things are comprised of autonomous, already individuated objects, or "atoms", and relations are represented by abstract links connecting these atoms. This classical view, which has been extensively studied in the philosophical literature (for an excellent discussion, see in particular Wittgenstein [414] and Mulligan [240]), is nominalist and forms the basis of logical atomism and set theory. In the gestaltic conception, by contrast, things and relations taken together constitute wholes, i.e., they are part of higher-order complex objects, holistic units or global configurations. First, things are separated and detached from their common background by a segmentation process (see previous chapter). Then, a mereological analysis can reveal what this "background" is made of, both in the spatial relations between the objects and in their morphological features. Here, relations between symbols are not taken for granted but *emerge together with the objects* through mereological segmentations and transformations.

One of the best presentations of this key idea, already promoted in the early beginnings of Gestalt theory and phenomenology, can be found in Husserl's 1907 *Ding und Raum* [161]. In §59 Husserl tries to understand how object configurations or "complexes" of objects are formed. Insisting that a typical feature of visual sensations is to constitute a field endowed with a spatial format, he explains that everything having a perceptual unity is rooted in the perceptual unity of the visual field (hence its temporal evolution, too). Individuated objects can only exist inasmuch as they are detached and profiled against their background. As a consequence, relations among objects are themselves profiled against the same background of spatial unity. In this sense, objects involved in relations can thus be conceived as being the parts of a *higher-order complex object*, the global configuration. The relations making a configuration of objects are therefore *mereological* and concern

the spatial order of parts at the core of the spatial whole.  $(p. 256)^1$ 

The main challenge is then to *categorize* these relations, and this constitutes a genuinely difficult problem.

From Husserl to contemporary research on AI, human-machine interaction and robotics, the interplay between perception and language in processing spatial relations has always been a central issue. Let us cite two examples, among many others. In his 1995 paper [151] "Coping with static and dynamic spatial relations" (see also [152]), Gerd Herzog stresses that

(the) interplay between visual perception and natural language in humanmachine interaction receives growing attention since it constitutes a prominent issue in many potential application areas. The aim of language-oriented AI research in this context is to achieve an operational form of referential semantics that reaches down to the sensoric level.

Going from visually accessible information to linguistic spatial descriptions is a crucial challenge that requires modeling spatial relations

<sup>1</sup> Independently of Husserl, it is well known in algebraic topology that the number of blobs  $X_i$  embedded in an ambient plane W is coded by the homotopy group of the complementary set  $W - \{X_i\}$ . This principle is widespread: the structure of a configuration X embedded in an ambient space E can be analyzed by looking at its complementary set W - X.

#### 1. INTRODUCTION

as intermediate conceptual units between linguistic and sensory levels to bridge from visual accessible three-dimensional geometric data to languageoriented representations.

In much the same vein, Nikolaos Mavridis and Deb Roy from the Cognitive Machine Group at the MIT Media Lab developed "grounded situation models for robots" [226] to bridge the gap between perception, action, and language.

Our long-term objective is to develop robots that engage in natural languagemediated cooperative tasks with humans. For example, the system can acquire parts of situations either by seeing them or by "imagining" them through descriptions given by the user: "There is a red ball at the left". These situations can later be used to create mental imagery, thus enabling bidirectional translation between perception and language.

The gestaltic conception calls into question the traditional perceptual vs. linguistic opposition. Visual perception is generally thought of as a faculty that is exclusively concerned with the shapes of objects, while language is supposed to be exclusively concerned with the abstract relations between those objects. Interestingly, this assumption, which rules over mainstream linguistics and the philosophy of language, is directly inherited from the classical gestaltic figure/background opposition. However, it starts losing its relevance when one takes into account the fact that the so-called perceptual "background" is a true figure in the eyes of language. As any figure, it is actively processed and organized by structuring principles—the grammatical elements (see below)—in a spatial and morphological way that is radically different from mere symbolic relationships. As we show in this chapter, the gestaltic conception of relations was remarkably revived and further developed by the recent trends in cognitive linguistics and cognitive grammars.

This conception is also very close to many psychological studies inspired by the fundamental works of Albert Michotte in the 1940's on the perception of causality. Against the traditional Humian view, Michotte [234] showed that apparently very high-level cognitive concepts such as *causality* are in fact deeply rooted in the automatic and hardwired algorithms of perception. As explained by Johan Wagemans ([399], p. 11):

Before Michotte, nearly all writers had treated causality as a high-level cognitive concept, and tended to think of the currency of perception in terms of only lower-level properties such as color, texture, and motion. Michotte, in this context, demonstrated that even seemingly "cognitive" properties such as causality may be processed in the visual system.

Michotte was inspired by Gestalt theory and claimed for instance that the visual "inferences" explaining Kanizsa amodal completion phenomena are automatisms of low-level visual processing (see, e.g., [235]). We claim that this central thesis is valid for geometric and kinematic relations in general.

In cognitive psychology, the categorization of spatial relations in preverbal conceptual thought has also been investigated extensively (see, e.g., Mandler [222] and McDonough [227]).

In this chapter we want to apply this perspective to the analysis of the topological contents of grammatical elements. We examine the link between the spatial structure of visual scenes and their linguistic descriptions. Based on a sampler of scene/description pairs, we attempt to reconstitute the toolbox of basic dynamical routines that underlie the grammatical and spatial organization of these descriptions. We then derive effective algorithms from these routines and propose a computational model explaining how the endless diversity of schematic visual scenes can be mapped to a small set of standard semantic labels. We also show how these algorithmic principles naturally relate to new developments in the neuroscience of vision and spiking neurons dynamics. In short, our proposal is that spatial and grammatical structures are related to each other through image transformations that rely on diffusion processes and wavelike neural activity.

To this aim we focus our attention on the remarkable work of Leonard Talmy (Chapter 1, Section 3.3), unsurpassed in its precision, richness of details and systematic ordering. Our intention is to use a few of the numerous examples given by Talmy as the foundation for a new model based on gestaltic visual routines.

In the remainder of this chapter, Section 2 applies an "active semantic" approach to the analysis of spatial grammatical elements and draws a link with visual processing routines. In Section 3 we discuss spatial invariance and the notion of "linguistic topology" in the context of mathematical geometry, which leads us to reaffirm the importance of transformation routines. Section 4 illustrates the specific difficulties raised by perceptual-semantic schemata with a detailed example, the English preposition "across". Modeling principles and algorithms are then introduced in Section 5, while numerical simulations on cellular automata are described in Section 6. Finally, we suggest a natural and plausible neural implementation of expansion-based morphology, based on waveform activity dynamics in spiking neural networks, in Section 7.

#### 1.2. Scope of this study

Before proceeding further, however, we wish to avoid any misunderstanding by pointing out the following important epistemological aspect: in this chapter, we restrict our analysis of prepositional contents to their *morphological and perceptually rooted kernel*. A full linguistic analysis of prepositional semantics would be of a much more complex nature and will not be addressed here.

Contemporary theories of semantics have shown that terms with spatiotemporal content are highly polysemous. For example, different uses of 'in' relate to qualitatively different spatial situations [150]:

(1) a. There is a cat  $in_1$  the house.

#### 1. INTRODUCTION

- b. There is a bird  $in_2$  the tree.
- c. There is a spoon  $in_3$  the cup.
- d. There is a crack  $in_4$  the vase.

Spatial prepositions typically form prototype-based radial categories (see Rosch [326]). Here, 'in<sub>1</sub>' represents the most prototypical "containment" schema, while 'in<sub>2</sub>' departs from it slightly by the texture and openness of the container. Many top-down knowledge-based inferences are also constantly operating on the data. For instance, the analysis of the 'in<sub>3</sub>' sentence must take into account the instrumental function of the spoon and its mereological decomposition into spoon head and spoon handle. The spoon handle is sticking out of the cup so the complete spoon is not physically inside the cup but only *metonymically*, via its head part.<sup>2</sup> In the subcategorical island 'in<sub>4</sub>', the container is a surface, not a volume, etc.

Another major departure from prototypical use comes from *metaphors*, a pervasive mechanism of cognitive organization and categorization rooted in the fundamental domains of space and time (see Lakoff [197], Johnson [173], Lakoff-Johnson [201], Talmy [374], [375]). Metaphorically, the preposition 'in' grammaticizes a generalized "containment" schema that can be applied not only to a great variety of concrete objects and scenes but also to highly abstract situations:

- (2) a. I am in<sub>A</sub> a committee.
  - b. I am  $in_B$  doubt.

These uses of 'in' pertain to a virtual concept of space generalized from real space, respectively:

(3) a. I am in<sub>5</sub> a crowd. b. I am in<sub>6</sub> water.

In these examples, ' $in_A$ ' metaphorically maps to an element of a discrete numerable set exemplified by ' $in_5$ ', while ' $in_B$ ' relates to an immersion into a continuous ambient substance such as ' $in_6$ '.

With these preliminary remarks in mind, we restrict the scope of this chapter to relatively homogeneous subcategories or *protosemantic* features, i.e., lowlevel compared to linguistic categories but high-level compared to local visual features. Our goal is to categorize scenes into elementary semantic *subclasses*, such as 'in<sub>1</sub>', or clusters of closely related subclasses, such as 'in<sub>1</sub>-in<sub>2</sub>'. We will not attempt to delineate the whole cultural complex formed by the English preposition 'in' by a single (and non-existent) universal.

However, even without additional semantic operators, we are still faced with the difficult problem of linking invariant semantic kernels to an infinite

<sup>2</sup> Langacker also emphasized this point In [207]: "As shown by *The flower in the vase*, where most of the flower protudes", the semantic content of 'in' is not only topological but also functional.

continuum of perceptual shape diversity. Even in their most typical, physical and concrete realizations taken independently, such as ' $in_1$ ',  $in_2$ ', etc., how can these invariant semantic kernels apply to perceptual data?

This challenge will be at the core of the present chapter and the remainder of this book. As mentioned above, contributing to its solution means focusing on a collection of low-level topological and dynamical routines that structure visual scenes in a gestaltic fashion. We are not implying, however, that top-down symbolic operations are not a full part of the cognitive act of language. After all, linguistic productions are made of symbolic utterances. Moreover, there must exist a higher-level control or "intentional" level of organization deciding which low-level routines should be applied to the scene and in what sequence. Yet, before these issues can be solved, the central interface between the lower perceptual and protosemantic levels of language must first be addressed.

## 2. Talmy's Gestalt semantics

As we have seen in Section 3.3 of Chapter 1, a fundamental thesis of Talmy is that language involves a system of organizational elements that structure the conceptual material in much the same way as visual elements structure the perceptual material.<sup>3</sup> He also shows that these elements are expressed *grammatically*, as opposed to lexically. The grammatical elements (prepositions 'in', 'above', etc.; declension and conjugation suffixes, and so on) form closed classes, i.e., restricted sets that contain relatively few members and that almost never expand in the course of a language's evolution. By contrast, the lexical elements ('bird', 'box', etc.) are extremely numerous and continuously produced, therefore creating open classes. Assuming that sentences evoke cognitive representations (CRs) in the listener's mind, Talmy claims that the grammatical elements provide the conceptual structure or "scaffolding" of these CRs, while the lexical elements.

Ronald Langacker [203] (see Section 3.2) also proposes that language essentially revolves around *categorization* at all levels of cognitive organization—as well among the things as among their relations (processes) or the temporal evolution of these relations (events). From there, the notion of *schematization* plays a central role in cognitive linguistics. Since its early Kantian origins, a "schema" is a procedure that can transform the content of a concept into a modus operandi for *constructing its referents*.

#### 2.1. Active semantics

Talmy's most compelling idea is that in a relational complex of objects where a figure—Langacker's trajector TR—is profiled against a ground—Langacker's landmark LM— the ground is not merely a passive frame for the figure, but the ground's geometry is also grammatically specified. In fact, the majority

<sup>3</sup> All major articles of Len Talmy published between 1972 and 1996 have been collected in the two volumes of his magnum opus *Toward a Cognitive Semantics* [374], [375].

of spatial grammatical elements treat the figure TR as a mere point while making various elaborate distinctions for the reference object LM. For example, consider the following three sentences:

- (4) a. The ball is in the box.
  - b. The ball is on the box.
  - c. The ball is 20ft away from the box.

Depending on the viewpoint, (4a) construes the box as a container volume, whereas (4b) focuses on the top surface of the box and (4c) reduces the box to a reference point. Thus, different grammatical elements scaffold reality in conceptually different ways. A classical theory of syntax, by contrast, would assimilate the box to a symbol irrespective of its actual geometrical shape or any other perceptual features, and assume that all spatial meaning is encapsulated in the abstract relationships 'in', 'on', and 'away from.'

Now, by giving a new significance to the object's geometry we are also challenged to find how this geometry varies with linguistic circumstances, i.e., how prepositions actually select and process certain morphological features from the perceptual data, while ignoring others. We propose that simple morphemes like the spatial elements 'in', 'above' or 'across', correspond in fact to visual processing algorithms that take perceptual shapes as input, transform them in a specific way, and deliver a semantic schema as output. We call these algorithms *morphodynamical routines* and the global structuring process *active semantics*.

If we take the tenets of cognitive linguistics seriously, we are therefore led to the following postulate: given an input perceptual scene that is already the outcome of low-level preprocessing tasks, such as boundary detection and object segmentation (see Chapter 2), language semantics is an active process that takes this scene to the next level of organization through a package of transformations or "morphodynamical" routines. These morphodynamical routines erase details and create new forms that evolve temporally.

- They enrich a visual scene with new *virtual* structures or *singularities* (e.g., influence zones, fictive boundaries, skeletons, intersection points, force fields) that did not originally belong to the scene but reveal the meaning it conveys.
- To this goal, they select from the scene its appropriate morphological features (surfaces, angles, segments, etc.) as the basic material to carry out the said transformations.

In summary, according to this program, semantic schemata literally trigger additional processing routines. We suggest below that these routines are akin to object-centered and diffusion-like non-linear dynamics.

## 2.2. Basic structuring schemata

In this section we recall and adapt a representative sample of Talmy's findings to illustrate the basic structuring operations performed by grammatical elements.

As explained above, our goal is to associate grammatical configurations with data-reducing transformation algorithms essential to scene categorization.

2.2.1. Magnitude invariance  $\rightarrow$  multi-scale analysis. We start with the linguistic properties of invariance and neutrality with respect to the detailed shape and dimensions of the perceptual data. One of Talmy's major claims ([374], p.28) is that:

the nature of [grammatical] structure is largely relativistic, topological, qualitative or approximative rather than absolute, Euclidean, quantitative or precisional.

We will discuss the "topological" quality of grammar in more detail in Section 3. For now, we retain the fact that grammatical elements are fundamentally scaleor metric-invariant. Compare, for example (Talmy [374], p. 25):

- (5) a. This speck is smaller than that speck.
  - b. This planet is smaller than that planet.

The above sentence pair shows that real distance and size do not play a role in the partitioning of space carried out by the deictic words 'this' and 'that'. The relative notions of proximity and remoteness conveyed by these elements are the same whether in millimeters or parsecs. This does not mean, however, that grammatical elements are insensitive to metric aspects but rather that they can uniformly handle similar metric configurations at vastly different *scales*. Multiscale processing therefore lies at the core of Gestalt semantics and, interestingly, as we have seen in Chapter 2, has also become a main area of research in computational vision (see, e.g., Mallat [219], Perona-Malik [257], Whitaker [401], Morel-Solimini [239]).

2.2.2. Morphology invariance. Numerous examples collected by Talmy show that grammatical elements are largely indifferent to the morphological details of objects or trajectories of moving objects, and that these details are mostly entrusted to lexical elements. This invariance of meaning points to the existence of underlying visual-semantic routines that perform a *drastic but targeted simplification* of the geometrical data.

Data-reducing routines fall into a few different categories, partly listed below. They are crucial to the faculty of scene *categorization*.

 $\diamond$  Bulk-neutrality  $\rightarrow$  skeleton routines

Another form of magnitude invariance appears in the selective rescaling of specific spatial dimensions. This is the case of thickness or bulk invariance. For example (Talmy [374], p. 31):

- (6) a. The caterpillar crawled up along the filament.
  - b. The caterpillar crawled up along the flagpole.
  - c. The caterpillar crawled up along the redwood tree.

These sentences show that 'along' is indifferent to the girth of LM and focuses exclusively on its main axis or *skeleton*, parallel to the direction of TR's trajectory. Like multi-scale analysis, skeleton transforms are widely used in computational vision and implemented using various algorithms (see the previous chapter). Studied by Blum [38] under the name "medial axis transform" and by others as "cut locus", "stick figures", "shock graphs" or "Voronoi diagrams" (see, e.g., Marr [224], Siddiqi et al. [345], Zhu-Yuille [419]), morphological skeletonization plays a crucial role in theories of perception and is considered a fundamental structuring principle of cognition by Leyton [212]. Skeletons indeed conveniently simplify and schematize shapes by getting rid of unnecessary details, while at the same time conserving their most important structural features. As we have seen in the previous chapter, there is also experimental evidence that the visual system effectively constructs the symmetry axis of shapes (see, e.g., Kimia [179] and Lee [209]).

 $\diamond$  Continuity-neutrality  $\rightarrow$  expansion routines Sentences such as:

- (7) a. The ball is in the box.b. The fruit is in the bowl.
  - c. The bird is in the cage.

are good examples of active-semantic analysis showing the neutrality of the preposition 'in' with respect to the morphological details of the container. We therefore propose that the active-semantic effect of 'in' is to trigger routines that transform individual objects and their composition in a scene in the following manner:

- The container LM ('box', 'bowl', 'cage') is closed by adding virtual structures to complete its missing parts: the 'bowl' becomes a sphere, the 'cage' a continuous surface of similar shape, etc. Generally, this type of *closing* process replaces objects by continuous spherical or blob-like occupation domains.
- The contained object TR ('ball', 'fruit', 'bird') expands, for example by a contour diffusion process, irrespective of its detailed shape until it collides with the boundaries of the closed LM. In fact, these two phases of the 'in' routine, the closing of LM and the expansion of TR, can run in parallel.
- Both TR and LM expand simultaneously until they encounter each other's boundaries. In this case, the expansion of LM naturally creates its own closing while at the same time providing an obstacle to the expansion of TR.

Therefore, we propose the following active-semantic definition of the prototypical "containment" schema (corresponding to the 'in<sub>1</sub>-in<sub>2</sub>' cluster of examples (1) and their closely related variants): A domain A lies 'in' a domain B if an isotropic expansion of A is stopped by the boundaries of B's closing or own isotropic expansion.

Although simple, the double expansion process that we propose for the "containment" schema is not trivial and leads us to introduce the following general principles of active-semantic morphodynamics:

(i) each object has a global 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* in an isotropic or anisotropic fashion. Image elements cooperate to propagate activity across the field and inhibit activity from other sources. Thus, the preposition 'in' involves an isotropic obstacle, whereas for example 'over' involves an obstacle in the lower half of the scene preventing a vertical expansion of the TR from reaching the bottom edge of the visual field.

In summary, within the general framework that all perceptual objects share the same surrounding space, we suggest that spatial interrelations between objects are inferred from the boundaries created by their territories (wave front collision curves and singular points on these curves) and the dynamical evolution of these boundaries. A directly measurable consequence of this definition is the fact that no TR-induced activity will be detected on the boundaries of the visual field. The final boundary between the domains is approximately equal to the skeleton of the *complementary space* between the objects, also called *skeleton by influence zones* or SKIZ (see Prewitt [316] and Serra [341]). In the case of the "containment" schema, the SKIZ surrounds TR and no region on the boundaries of the image will be encompassed in TR's expanded domain. The sheer absence of TR activity on the boundaries is a robust feature that contributes to categorize the scene as 'in'. It illustrates a typical morphodynamical routine at the basis of our perceptual-semantic classifier.

 $\diamond$  Shape-neutrality  $\rightarrow$  singularities

Other examples by Talmy clearly show that salient aspects of shapes are simply not taken into account by grammatical elements. In the following sentences (Talmy [374], p. 27):

- (8) a. I zigzagged through the woods.
  - b. I circled through the woods.
  - c. I dashed through the woods.

the actual trajectory of the moving object TR (the first person), whether made of segments, curves or a single straight line, ultimately connects two opposite (and possibly virtual) sides of the extended domain LM (the woods). In our active-semantic view, this shows that the element 'through' creates a drastic schematization of TR in two steps:

#### 3. WHAT IS "COGNITIVE TOPOLOGY"?

- the trajectory is "convexified" or "tubified", again by outward expansion (yet, here, to a limited extent) and becomes what is called in geometry a *tubular neighborhood*; then
- the main axis of this tube is extracted through a *skeletonization* process that is basically a reverse inward expansion eroding the object down to its medial symmetry axis.

At the same time, it does the same with LM: the texture is first erased, then the domain is skeletonized. In short, the grammatical perspective enforced by the English preposition 'through' reduces the detailed original trajectory of TR and texture of LM to mere fluctuations around their medial axes. Finally, the characteristic feature of the 'through' schema lies at the intersection of the two skeletons.

As already mentioned above, skeleton transforms are widely used in computational vision and implemented using various algorithms. They offer a convenient way to simplify and schematize shapes by getting rid of unnecessary morphological details, while at the same time conserving their most important structural features. For this reason they are also well suited to the analysis of relations. We will further develop this point below (see Section 4).

In this example, the spatial relation between the objects is inferred from virtual *singularities* in the boundaries of the objects' territories. These singularities and their dynamical evolution are important clues that constitute the characteristic "signature" of the spatial relationship (Petitot [261]). Transformation routines considerably reduce the dimensionality of the input space, literally "boiling down" the input images to a few critical features. A key idea is that singularities encode a lot of the image's geometrical information in an extremely compact and localized manner.

In summary, after the morphodynamical routines have transformed the scene according to the expansion principles (i) and (ii), several types of characteristic features can be detected to contribute to the final categorization:

- (iii) presence or absence of activity on the boundaries,
- (iv) intersection of skeletons,
- (v) singularities in the SKIZ boundary.

Our active-semantic approach proposes a link between the "things, relations, events" trilogy of cognitive grammar (Langacker [203]) and the "domains, singularities, bifurcations" trilogy of morphodynamical visual processing (Petitot [261]). We further suggest in Section 7 that the brain might rely on dynamical activity patterns of this kind to perform invariant spatial categorization.

# 3. What is "cognitive topology"?

The core invariance of spatial meaning reviewed above is sometimes referred to as the "linguistic form of topology" (Talmy [374]), or "cognitive topology" (CT) (Lakoff [199]), and is often compared to mathematical topology (MT).

Section 2.2 showed that, just like MT, CT is magnitude- and shape-neutral. Yet, CT is also more abstract and less abstract than MT at the same time (Talmy [374], p. 30):

- In some areas, CT has a greater power of generalization than MT. For example, the invariant "containment" schema extracted from 'in the box', 'in the bowl' or 'in the cage' (see 2.2.2) shows that incomplete or disconnected parts can be equivalent to a single continuous surface through a closing process.
- In other areas, however, CT also preserves metric ratios and limits distortions in a stricter sense than MT. For example, the lexical elements 'cup' and 'plate' have distinct spatial semantic properties and uses, although a plate is nothing more than a flattened version of a cup (they are homeomorphic in the mathematical sense). As we will later see in Section 4, the grammatical element 'across' also exhibits subtle metric constraints in its applicability.

In reality, because of all these discrepancies, CT has actually little to do with true MT. In the mathematical sense, topology refers only to a certain level of structural description of spatial objects. Let us give a brief reminder of this topic.

In mathematics, structural levels go from the least constrained (sets) to the most constrained (Euclidean metric) and each level is associated to a particular class of mappings or structure morphisms preserving that level:

- Level 0 corresponds to pure *set* structures. Points are independent from each other and mappings can be any set applications.
- Level 1 is the *topological* level. Points are "glued together" through qualitative neighborhood relationships and mappings are continuous applications, which can be for example fractal.
- Level 2 is the *differentiable* level. Objects are "smooth" and can be approximated by tangent linear objects. Mappings are infinitely differentiable. Fractals are typical examples of objects from Level 1 that do not belong to Level 2.
- Level 3 is the *conformal* level and corresponds to the existence of "complex" structures, in the sense of complex numbers. Mappings are holomorphic, which means in the 2D case that they preserve angles.
- Level 4 is the *metric* level and introduces the concept of distance. Several subcategories can be distinguished here: In Riemannian spaces, metric is non-homogeneous (it is local and can vary from point to point); in Euclidean or non-Euclidean spaces (hyperbolic, Minkowskian, etc.) metric is homogeneous and mappings are isometric transformations.
- Level 5 is both metric and *linear*. This is the level of vectorial spaces with a norm (Euclidean spaces, Hilbert spaces).

We want to address again the question of a cognitive topology within this mathematical context. Levels 2 and 3, which lie between the malleable topological level and the rigid metrical level, are both "soft" in the sense that they allow stretching, and are at the same time more constrained than Level 1. At first sight, these qualities would make Levels 2 and 3 good candidates for the sought-after cognitive level of representation of CT. However, we think that this view is misleading. In our sense, if there is such a thing as a cognitive topology, it does not correspond to any of the levels (or intermediate levels) in the above mathematical hierarchy.

## 3.1. Convexification

In reality, the active semantics conception presented in Section 2.1 leads us to consider *inter-level* transformations, i.e., processes starting at one level and going up or down the hierarchy. For example, a *convexification* routine, which makes objects smoother and "rounder", is a true schematization in the sense that it discards a great amount of the objects' metric properties and yields only a small number of blob-like shape categories. To the limit of spherical convexification, all objects become equivalent to a unique 'sphere' class. This makes convexification appear to go down the level hierarchy, towards the "soft" topological levels. Yet, at the same time, the convex shapes it creates are more "rigid" than the original objects because, precisely, mappings acting within these shape classes are constrained to preserve convexity. Therefore, convex shapes somehow also belong to the higher metric levels.

This example illustrates the subtle and complex interplay between a process of impoverishment of details and a process of rigidification of structures. On the one hand, CT is "neutral" toward certain structural aspects (magnitude, continuity, shape, bulk, etc., see 2.2), which seems to imply that it lies in the lower range of the hierarchy of mathematical levels. However, on the other hand, active-semantic or morphodynamical routines essentially transform objects into *new* objects that are categorical prototypes and, precisely because they are prototypical, by nature more rigid and metrically constrained. For example, by transforming the topological and differentiable "amoeboid" forms of levels 1 and 2 into regular, prototypical spheres we are in fact going up the level hierarchy, towards the metric levels.

In short, one could say that schematization and categorization replace "soft complex" forms with "rigid simple" ones. Here lies the puzzling apparent paradox of cognitive topology.

#### 3.2. Skeletonization

Another remarkable example of this phenomena is *skeletonization* (seen in Section 2.2.2). Like convexification, skeletonization performs a drastic simplification of objects, yet it is also highly sensitive to their metric details. An interesting problem is then to find out what transformations preserve the topology

of the skeleton. If we denote by Sk(X) the skeleton of a shape X, the skeletonbased relation of equivalence  $\cong$  between two shapes

(1) 
$$X \cong Y \iff \operatorname{Sk}(X) \simeq \operatorname{Sk}(Y)$$

where  $\simeq$ , is much weaker than an isometry. Yet, it is also much more rigid than conformal isomorphisms and, a fortiori, diffeomorphisms and homeomorphisms.

In summary, active semantics is first and foremost a matter of processes, transformations, representations and re-coding. It does not correspond to any of the elementary levels of mathematical structures. As we will see below, it nevertheless corresponds to a family of operations that can be called morphological and it is in that sense that one may, somehow incorrectly, speak of a "morphological level" for cognitive topology.

**Thesis:** What is generally called "topology" in Gestalt semantics and cognitive grammar actually corresponds to the action of morphological operators.

# 4. Operations on schemata: the 'across' puzzle

In order to illustrate the richness and difficulty of the problem raised by perceptual-semantic schemata, even leaving out their metonymic or metaphorical extensions and restricting ourselves to invariant perceptual kernels (see Section 1.2), we want to study here the example of the preposition 'across' in greater detail.

#### 4.1. Invariant of transversality

The main geometrical concept underlying all uses of 'across' is the concept of "transversality", It is a natural and intuitive concept that was clarified formally only recently within the framework of modern differential geometry, in particular by the works of Hassler Whitney and René Thom. Let us start with a few simple considerations:

- Intuitively, two distinct lines  $L_1$  and  $L_2$  in the plane  $\mathbb{R}^2$  are transverse if they intersect, i.e.,  $L_1 \neq L_2$  and  $L_1 \cap L_2 \neq \emptyset$ . This is a stable or "generic" situation in the sense that small perturbations cannot make transverse lines parallel.
- By extension, two (smooth) curves  $C_1$  and  $C_2$  in the plane  $\mathbb{R}^2$  are transverse if they intersect at points where their tangent lines  $L_1$  and  $L_2$  are transverse.
- Two lines  $L_1$  and  $L_2$  in the space  $\mathbb{R}^3$  are never transverse when they are coplanar, because this would not be a generic situation: any infinitesimal perturbation could bring them apart.
- On the other hand, a line L and a plane P such that  $L \not\subset P$  in the space  $\mathbb{R}^3$  are transverse if they intersect.

- By extension, a (smooth) curve C and a (smooth) surface S in the space  $\mathbb{R}^3$  are transverse if they intersect at points where their tangents, respectively a line L and a plane P, are transverse in the previous sense.
- Two distinct planes  $P_1$  and  $P_2$  of the space  $\mathbb{R}^3$  are transverse if they intersect.
- By extension, two (smooth) surfaces  $S_1$  and  $S_2$  are transverse at one of their intersection points if their tangent planes  $P_1$  and  $P_2$  at that point are transverse.

Extracting an invariant kernel from these various instances of transversality, one comes up with the following formal definition:

• Two (smooth) subspaces U and V of an ambient space M, for example  $\mathbb{R}^2$  or  $\mathbb{R}^3$ , are said to be transverse at one of their intersection points x if their tangent spaces at  $x, T_x(U)$  and  $T_x(V)$ , generate the whole tangent space of M at x, i.e., if  $T_x(U) + T_x(V) = T_x(M)$ .

#### 4.2. Variants of transversality

Several archetypical schemata of transversality can be developed around this general conceptual kernel. One of them is the rather simple "crossing" schema, in the sense of road-crossing, and involves two 1D curves intersecting in a plane. Another schema, more complex, corresponds to the English preposition 'across' and will be examined in the rest of this section. It involves a domain D (the LM) delimited by boundaries, and a path C (the TR) traversing this domain. More precisely, the 'across' schema is in fact characterized by a double transversality and can be tentatively described as follows:

Across 1: A path C is going across a domain D if (a) it enters D at a point  $x_1$  where it intersects the boundary  $\partial D$  of D transversally, (b) it lies inside D, and (c) it exits D at a point  $x_2$  on an "opposite" segment of  $\partial D$  where it intersects  $\partial D$  transversally again

Therefore, the particular perceptual-semantic schema conveyed by 'across' includes a notion of transversality between a 1D curve and a 2D domain of the plane, itself defined as a double transversality between the curve and the domain's boundary. The rigid prototypical form of this schema corresponds to a line intersecting a rectangle perpendicularly (Figure 1).

#### 4.3. Plasticity of perceptual-semantic schemata

Talmy identified a great number of different uses of the 'across' schema. These examples are typical of the high degree of plasticity displayed by perceptosemantic schemata.

4.3.1. Topological plasticity. First, the path C and the boundary  $\partial D$  can be considerably deformed without affecting the 'across' property, as long as the three criteria in the above **Across 1** definition are preserved (Figure 2).



FIGURE 1. The 'across' prototype: a line intersecting a rectangle perpendicularly



FIGURE 2. The 'across' prototype is plastic and can be greatly deformed up to a certain limit without losing its meaning.



FIGURE 3. A bad instance of 'across': the exit point  $x_2$  is not on the boundary "opposite" to the entry point  $x_1$ 

However, this necessary condition is not sufficient. It is easy to show that, although they can be deformed to a great extent, neither C nor  $\partial D$  can be completely *arbitrarily* deformed as they would be in a purely topological sense, while preserving the 'across' property. In fact, identifying these limits precisely constitutes the core difficulty of the 'across' puzzle, as we will see below. For example, Figure 3 is not a good instance of 'across', because  $x_2$  is not on the boundary "opposite" to  $x_1$ —but, again, what is an "opposite" boundary?

4.3.2. The concrete/virtual dialectics. Talmy gives numerous examples showing that the schema does not need to be complete and that one or more of its

components can be virtual or missing. They span an entire range between two extreme cases:

(9) a. The tumble weeds are swept across the prairie.b. The snail crawled across the car.

In (9a) the boundary  $\partial D$  of the 'prairie' LM domain is virtual, and therefore  $x_1$  and  $x_2$  too, whereas in (9b) it is the path C of the 'snail' TR that is virtual.

4.3.3. The 2D/3D ambiguity. Talmy also gives adaptive examples of applicability of schemata. The following pair of sentences:

(10) a. He flew across/over the plateau.b. He walked across/through the wheat field.

In (10a), 'across' takes the sense of 'over': the 'plateau' domain is 2D and the condition of transversality applies to two curves in a plane, the path C and the boundary  $\partial D$ . In (10b), by contrast, 'across' takes the sense of 'through': the 'wheat field' domain is now a 3D volume and its boundary a 2D surface. Here, transversality occurs between a curve C and a surface in a 3D ambient space.

4.3.4. *Constraints on distortion.* We just saw that there were additional constraints to the plasticity of a schema and that it cannot be of a purely topological nature, allowing free-range deformations. To further appreciate the problem in all its subtlety, let us take the example of a lake:

- A path that penetrates the domain of the lake only slightly, within its outer margin and then follows the shore to eventually exit on the same side (Figure 4a) does not really constitute a situation of transversality. It is rather an approximatively "tangential" situation.
- A path that penetrates the domain of the lake, goes almost all the way through to the other side but then turns around and comes back to exit on the entrance side poses a more delicate problem (Figure 4b). One can argue that it is not a good case of transversality either, because the second crossing  $x_2$  is not on the "other side" of  $x_1$ —which again raises the question of a definition of the "other side".

These difficulties are characteristic of the morphological nature of perception and cannot be simply resolved by using some kind of metric template matching (levels 4 and 5 of the mathematical hierarchy evoked in Section 3) or, on the other end of the spectrum, by using free-range topological deformations (levels 1 and 2). On the one hand, template matching would be an unpractical task because of the sheer number of templates necessary to cover the combinatorial complexity of typical and subtypical configurations. On the other hand, a more flexible template matching allowing greater deformations would in fact accept bad configurations and yield a lot of false positives.



FIGURE 4. The difficulty of the 'across' puzzle and a proposal for a solution. (a) A path that is too "tangential" to the boundary cannot be described by 'across'. (b) A path that is sufficiently transverse to the boundary but exits on the "same" side cannot be described by 'across' either. (c) To be a good example of 'across', a path must intersect transversally the medial axis (cut locus) an odd number of times (see below).

#### 4.4. Virtual structures

To find a way out of the misleading topological vs. metric dilemma, the solution we propose involves creating virtual structures that embody metric information without being themselves metric. The 'across' schema contains an implicit partitioning of the domain D into 'this side' and 'that side' (see 2.2.1), equivalent to the partitioning performed by 'here' and 'there'. This qualitative difference can be geometrically implemented by a virtual intermediate boundary  $C_D$  approximately located in the middle of the domain. Before the boundary  $C_D$ , we are on 'this side' of the domain and after  $C_D$  we are on 'that side.'

An interesting property of the virtual boundary  $C_D$  is that it is both of a metric nature, since it lies in the "middle" of D, and at the same time of a qualitative nature, since its only purpose is to divide D into two subdomains. The position of  $C_D$  is quantitatively precise (at maximal distance from the borders of D) but its function is really to discretize the underlying continuum. This is similar to the origin point on a coordinate axis: both precisely located and qualitatively arbitrary, present only to separate two half-lines.

The virtual structure  $C_D$  we need naturally corresponds to the "cut locus" (or "skeleton", "symmetry axis", etc.) already mentioned in Section 2.2 about the magnitude and morphology invariance of grammatical elements. A simple skeletonization of the domain D solves the problem rather easily, albeit partially, and the above criterion can be modified in the following manner:

Across 2: A path C is going across a domain D if (a) it lies inside D, (b) is transverse to the cut locus  $C_D$  of the domain in one or an odd number of points, and (c) intersects the boundary  $\partial D$  of the domain in two points:  $x_1$  on one side of  $C_D$  when C enters D, and  $x_2$  on the other side of  $C_D$  when C exits D (Figure 4c).



FIGURE 5. Ambiguity between 'across' and 'along.' Left: 'across' the swimming pool lane. Right: 'along' the swimming pool lane.

Therefore, we find again the pure invariant of transversality applied this time to the intersection between the cut locus  $C_D$  and the path C. The transversality archetype also operates on transformed shapes, giving the schema all its plasticity. For example, the path C can zigzag around the cut locus as long as the net result is a crossing from one side to the other.

In summary, this example clearly illustrates the fact that obtaining the right type of constraints, in order to characterize perceptual-semantic schemata, requires taking into account additional virtual structures.

4.4.1. Across vs. along. Another compelling demonstration of the idea that morphological analyzers such as the cut locus can help solve cognitive-topological puzzles is given by the following two sentences (Figure 5):

- (11) a. I swam across the swimming pool lane.
  - b. I swam along the swimming pool lane.

Looking at Figure 5, the conditions for using 'across' seem to be satisfied in both cases according to the first definition **Across 1** above. Yet, in the situation depicted by (11b) 'across' is obviously incorrect in English and should be replaced by 'along'. Why is this the case? To resolve this issue, one could try to amend **Across 1** by introducing an additional *ad hoc* metric constraint that can discriminate between (11a) and (11b). This constraint could rely on the *ratio* of the lane edges with respect to the direction of the path: the edge parallel to the path should be shorter than the edge perpendicular to the path, i.e., their ratio should be less than 1. But this kind of metric criterion is not satisfactory because it is in fact too metrical (Euclidean). It is applicable only to a small number of well-defined configurations, i.e., regular rectangle pool lanes in which the swimmer's path can be "parallel" to one edge and "perpendicular" to the other. It also loses the link with the topological invariant of transversality.



FIGURE 6. A solution to the 'across'/'along' ambiguity. Left: 'across' applies because the path is transverse to the cut locus. Right: 'along' applies because the path is not transverse to the cut locus (here, parallel).

Instead, by introducing the cut locus  $C_D$  of the domain and applying the modified definition **Across 2**, the discrimination between (11a) and (11b) becomes natural and immediate. Whereas in case (11a) the path is *transverse* to  $C_D$ , in case (11b) it is not (here, parallel to  $C_D$ ) (Figure 6).

However, this rather simple and elegant criterion also has limits in special cases: for example, if the path in (11a) was too close to one of the short edges of the pool lane, it would not intersect  $C_D$  anymore. Nevertheless, we believe that this kind of criteria involving basic virtual structures are able to bridge the gap between perception and semantics.

## 4.5. Other examples of virtual structures: fictive motion

The introduction of virtual, fictive, or subjective structures is crucial to Gestalt semantics. Len Talmy uncovered and categorized a wealth of linguistic instances in which fictive structures are grammatically embodied in language. In particular, his unified system of "fictive motion" (Talmy [373], [374], chap.2), a semantic phenomenon creating a sense of motion without physical realization, displays one of the most overt types of such structures. To illustrate this point, we give here a short sample of fictive motion categories identified by Talmy.

• Emanation paths. A finger pointing at an object draws a virtual line through space linking it to the object, and beyond:

(12) I pointed him toward/past/away from the lobby.

• Shadow paths. An object's shadow is perceived as moving from the object to a background surface:

#### 5. MODELING PRINCIPLES AND ALGORITHMS

(13) a. The tree cast its shadow down into the valley.b. The pillar's shadow fell onto the wall.

• Virtual processes. In many cases, a static situation is construed as the result of a virtual dynamical process:

(14) a. The palm trees clustered together around the oasis.b. Mountains are scattered all over this region.

• Virtual fronts. Any object defines a "front" comprising a virtual plane and a path emanating perpendicularly from this plane. This virtual structure is present in demonstrative paths, sensory paths, and radiation paths.

# 5. Modeling principles and algorithms

## 5.1. Gestalt computation

The family of problems that we examined in the previous sections constitute Gestalt-type problems in the strong technical sense of the term. They involve global structures, virtual organizer elements, dialectics between the quantitative and the qualitative, the metric and the categorical, along with properties of topological and morphological plasticity. They amplify the classical Gestalt phenomena (such as virtual contours or the *Prägnanz* of right shapes). At the algorithmic level, they also raise an arduous local/global dilemma: how to strike a balance between global Gestalts on the one hand, and local computation on the other hand.

The main issue is the following: what kind of computational procedures are able to generate (through a "scanning" process) the aforementioned virtual organizing structures? The core difficulty is that global (holistic) elements containing infinite information have to be processed by purely *local* algorithms only using informationally finite procedures. From this perspective, cognitive grammars are not satisfactory: they rely on rather mysterious and holistic properties of vision without attempting to create explicit models and work out algorithms that could explain these properties.

Moreover, those local procedures must derive from principles that are general enough to successfully explain how any unsupervised *learning* of the closed grammatical classes, with all their fascinating perceptual-semantic nuances, can even be possible.

## 5.2. Spreading activation

In fact, the structuring operations that we sketched in Section 2 all point toward a specific type of process. To be able to generate global structures using local mechanisms, the only possible solution is to trigger *diffusion-propagation* or *expansion processes* from the salient features of the original image, in particular the contours. These contours become active (excitatory or inhibitory) and trigger "activity waves". In turn, these waves create wave front singularities and, with multiple sources, wave front collisions.

To recapitulate, the routines that we gathered in our exploration of numerous linguistic examples are: convexification, tubification, skeletonization, cut-locus, triggering contours and obstacle contours, etc. Drawing from these observations, we propose the two following key ideas toward the foundation of an active morphological semantics:

**Key idea 1**: Contours are active elements that either trigger processes of diffusion-propagation activity, or inhibit such processes coming from other sources.

Key idea 2: Diffusion-propagation processes generate in turn new geometrical entities, in particular singularities, that act as virtual structuring elements.

## 5.3. Links with other works

5.3.1. Terry Regier. In his 1988 article [320] Recognizing Image-schemata Using Programmable Networks (see also Harris [144]), Terry Regier also used a contour diffusion routine for analyzing the cognitive (perceptual-semantic) content of prepositions such as 'in', 'above', etc. He starts from the hypothesis that

recognizing an image-schema for a given preposition amounts to recognizing that the particular spatial configuration named by the preposition holds between the entities focused on in the current field of view. (p. 315)

In order to organize the scene, he uses Ullman's local visual routines, and then algorithms of "Bounded Spreading Activation", that is, contour diffusion.

The basic idea is that activation spreads out from the copies of object A that were placed in the working image nets, and stops when it encounters units that correspond to the borders of the object in the *B*-net. (p. 318)

After this first model, Terry Regier refined his theory and, with Laura Carlson [51], introduced the role of *attention*: when processing spatial relations, geometrical information is mixed through attention with functional information provided by the mereological decomposition of objects into functional parts.<sup>4</sup> Taking into account mereology allowed him to develop a finer model, the AVS (Attentional Vector-Sum) model [322] in which projective prepositions such as 'above' are

grounded in the process of attention and a vector-sum coding of overall direction.

The idea is to look at the functional parts  $LM_i$  of the LM, to weight the subrelations  $LM_i/TR$  by the attentional focus, and to take the weighted sum.

<sup>4</sup> See above, Section 1.2, the example of the spoon in the cup. It is the functional part "spoon's head" (not the handle) that is in the cup.

5.3.2. Seibert and Waxman. In their article [338] 'Spreading Activation Layers, Visual Saccades, and Invariant Representations for Neural Pattern Recognition Systems", Michael Seibert and Allen Waxman have built a model, the NADEL model—"Neural Analog Diffusion-Enhancement Layer"—for bottom-up pattern recognition. They have shown that a spreading activation network—a 2D diffusion—including a detection of local maxima

can quickly perform a large number of early vision tasks (...) in a completely data-driven manner. (p. 9)

One of the main interest of their work is to have introduced a feed-back of the detection of critical points (maxima) on the diffusion itself.

This feed back pattern provides a short-term memory reverberation between the levels, sharper peaks, and automatic gain control. (p. 14)

# 5.4. Morphological algorithms

The above hypotheses about Gestalt semantics must be empirically verified by computational synthesis. However, practical technical difficulties arise when trying to implement the standard differential equations of non-linear systems, such as reaction-diffusion or traveling waves. These equations require a significant amount of computing power and parameter tuning, therefore they are not the fastest way to model the elementary mechanisms of expansions and obstacles at work in perceptual semantics. We will come back to "dynamical systems" type of implementations in Sections 6 and 7, through cellular automata and spiking neurons, but for now we adopt a "functionalist" approach (still within the framework of dynamical models) and model these physicalist algorithms with more abstract counterparts found in *mathematical morphology* (MM). This domain of mathematics was developed in the 1970's at Ecole des Mines de Fontainebleau, France, by Georges Matheron and his colleagues Jean Serra [341], M. Schmitt, J. Mattioli and others. MM is a branch of image processing that focuses on the geometric structure of shapes and textures. We present here a review of the main concepts developed by MM (for an introduction see also Dougherty [87]) and draw a link to Gestalt semantics. Next, in Section 6, we present a series of numerical simulations of MM-based active semantics implemented in *cellular automata* (CA).

Lastly, in Section 7, we explore the deeper, physicalist, microstructure of active semantics. Specifically, we will establish a link between MM and *spiking neural networks*, which can play the role of excitable media supporting complex waveform dynamics. Dynamical events of an intrinsically *spatio-temporal* nature that are critical for the emergence of virtual structures and singularities cannot be captured by MM and can be realized only at this finer level of resolution.

5.4.1. Morphological operations on binary images. The main idea is to analyze images by using structuring elements and Boolean set operations. Let X be



FIGURE 7. To illustrate the algorithms of mathematical morphology we use this picture of the celebrated Scythian Deer from the State Hermitage Museum in Saint Petersburg, Russia. Left: the original image. Right: its binary outline.

a 2D binary image defined on a window W of  $\mathbb{R}^2$  centered around the origin O. X is defined by a characteristic function  $\varphi_X : W \to \{0,1\}$ , such that the figure corresponds to the inverse image  $X = \varphi_X^{-1}(1)$  and the background against which it is profiled to  $\varphi_X^{-1}(0)$  (Figure 7).

We now introduce a basic structuring element B in the form of a small disk of radius r centered around O. Element B should be viewed not just as a geometric element but also a dynamical element resulting from a "binary diffusion" of O during a characteristic period of time. However, an important difference between binary diffusion and real PDE-based diffusion is that the element B remains black and white without any "smoothing" of its boundaries. It remains discrete on the range of image values and is defined by its characteristic function  $\varphi_B$ :  $a \in B \iff \varphi_B(a) = 1$ . Binary diffusion is a simplified version of real diffusion especially well suited to binary morphological analysis.

 $\diamond$  Dilation

Considering each point of the image X as an elementary source of binary diffusion, we can define the *dilation* of X by B as follows (Figure 8):

(2) 
$$X \oplus B = \{u \mid u = x + a, x \in X, a \in B\}.$$

The effect of the dilation of X by disk B is to regularize the image. It connects components inside X that are sufficiently close, fills up small gaps ("close" and "small" with respect to B's size) and globally increases the linear size of X by B's radius, r. The exact result depends of course on B's shape and size, and on the various ways of implementing this continuous formalism in a discrete lattice, e.g., the type of neighborhood (4-neighbor square, 8-neighbor square, 6-neighbor hexagonal, etc.).

If we denote by  $B_u$  the translation of B by u, we get

(3) 
$$B_u = \{v \mid v = a + u, a \in B\},\$$



FIGURE 8. A dilation of the deer image after 20 steps. Holes are gradually filled.

and we can then define the operation of dilation as follows:

(4) 
$$X \oplus B = \{ u \,|\, B_u \cap X \neq \emptyset \} = \bigcup_{x \in X} B_x = \bigcup_{a \in B} X_a.$$

Morphological dilation has interesting algebraic properties, such as commutativity and associativity:

(5) 
$$\begin{array}{rcl} X \oplus B &=& B \oplus X \\ (X \oplus B) \oplus B' &=& X \oplus (B \oplus B'), \end{array}$$

and partial distributivity with respect to set operations:

(6) 
$$X \oplus (B \cup B') = (X \oplus B) \cup (X \oplus B')$$
$$X \oplus (B \cap B') \subset (X \oplus B) \cap (X \oplus B').$$

It is also an increasing function:

(7) 
$$X \subset X' \Rightarrow (X \oplus B) \subset (X' \oplus B),$$

and is translation invariant:

(8) 
$$X_u \oplus B = (X \oplus B)_u.$$

Starting from dilation, we can now define other basic morphological routines, such as erosion, closing and opening.

 $\diamond \underline{\text{Erosion}}$ 

The erosion of X by B is the dual operation of dilation by set complementation. If the complement of X is denoted by

(9) 
$$X^c = W - X = \{ u \mid u \notin X \},$$

then erosion corresponds to the dilation of the complement in the following manner (Figure 9):

(10) 
$$X \ominus B = (X^c \oplus (-B))^c.$$

Image erosion obviously has the inverse effect of image dilation: it disconnects components linked by narrow strips, erases small spots (again, "narrow" and



FIGURE 9. Two stages (20 and 50 steps) during the erosion of the deer image. Holes gradually disappear, while the shape contracts toward its skeleton.

"small" with respect to B's size) and globally decreases the size of X by B's radius, r.

As in (4) it is easy to deduce alternative definitions of erosion based on translated sets:

 $a \in B$ 

(11) 
$$X \ominus B = \{u \mid u + a \in X, a \in B\}$$
$$= \{u \mid B_u \subset X\}$$

$$(12) \qquad \qquad = \bigcap X_{-a},$$

including the following algebraic property:

(13) 
$$X \ominus (B \oplus B') = (X \ominus B) \ominus B',$$

and set-theoretic properties:

(14) 
$$\begin{array}{rcl} X \ominus (B \cap B') &\supset & (X \ominus B) \cup (X \ominus B') \\ X \ominus (B \cup B') &= & (X \ominus B) \cap (X \ominus B'). \end{array}$$

Like dilation, erosion is an increasing function:

(15) 
$$X \subset X' \Rightarrow (X \ominus B) \subset (X' \ominus B),$$

and is translation invariant:

(16) 
$$\begin{aligned} X_u \ominus B &= (X \ominus B)_u \\ X \ominus B_u &= (X \ominus B)_{-u}. \end{aligned}$$

 $\diamond$  Closing

The closing of X by B consists of a dilation followed by an erosion:

(17) 
$$X \bullet B = (X \oplus B) \ominus B$$

Like dilation, closing connects segments and fills-in gaps but it also renormalizes the size of X and therefore has a global effect of regularizing or "smoothing" the image without changing its size (Figure 10).

Closing is a natural match for continuity-neutral schemata like the preposition 'in' (see 2.2.2): it transforms a group of disconnected segments into a



FIGURE 10. Closing the deer image (30 steps).



FIGURE 11. Opening the deer image (30 steps).

continuous curve (or surface in 3D). Note that closing is idempotent, i.e., it does not further modify an already closed shape:

(18) 
$$(X \bullet B) \bullet B = X \bullet B$$

 $\diamond$  Opening

The opening of X by B is the dual operation of closing and consists of an erosion followed by a dilation (Figure 11):

(19) 
$$X \circ B = (X \ominus B) \oplus B.$$

Just like erosion corresponds to the dilation of the complement, opening corresponds to the closing of the complement:

(20) 
$$X \circ B = (X^c \bullet (-B))^c,$$

and, like closing, is also idempotent:

(21) 
$$(X \circ B) \circ B = X \circ B.$$

5.4.2. Skeletonization of binary images.  $\diamond$  Medial axis skeleton

3. RELATIONS



FIGURE 12. The morphological skeleton of the Scythian deer.

Using morphological routines it becomes easy to compute the skeleton Sk(X) of a shape X, an otherwise complicated task when using partial differential equations. As we saw in Chapter 2, Sk(X) is made of the centers of all the maximal disks included in X (Figure 12).

Let us formalize this notion. For any subset X, we denote by X its topological interior (the union of all open sets contained in X) and  $\overline{X}$  its topological closure (the intersection of all closed sets containing X).<sup>5</sup> To avoid pathological cases, we assume in the remainder of this section that X is already an *open* set, is connected, has an arc-wise differentiable boundary  $\partial X$ , and that its topological closure  $\overline{X}$  is compact. We denote by B(x, r) the open disk of center x and radius r, and  $\overline{B}(x, r)$  its topological closure:

(22) 
$$B(x,r) = \{u \mid d(x,u) < r\} \\ \overline{B}(x,r) = \{u \mid d(x,u) \le r\},\$$

where d is the distance in the ambient space. The disks centered around the origin are denoted B(r) = B(0, r) and  $\overline{B}(r) = \overline{B}(0, r)$ . A disk B(x, r) is said to be maximal in X if it verifies

(23) 
$$\forall x' \forall r' B(x,r) \subset B(x',r') \subset X \Rightarrow x' = x \text{ and } r' = r,$$

i.e., if no larger disk can be inserted between B(x,r) and the boundary of X. If we now consider the erosion transformation of X by B(r), we get:

(24) 
$$\begin{aligned} X(-r) &= X \ominus \overline{B}(r) \\ \overline{X}(-r) &= X \ominus B(r) \end{aligned}$$

<sup>5</sup>  $\overset{\circ}{X}$  and  $\overline{X}$  are classical topological objects and should not be confused with the morphological operations of opening,  $X \circ B$ , and closing,  $X \bullet B$ . Reminder in 1D: if X is the interval [-1, 1), including -1 but not 1, then  $\overline{X} = [-1, 1]$  and  $\overset{\circ}{X} = (-1, 1)$ .

it is easy to prove that a disk B(x,r) is maximal in X if and only if x belongs to  $\overline{X}(-r)$  but not to any morphological opening of  $\overline{X}(-r)$  by any closed disk  $\overline{B'}(\varepsilon)$ . In other terms, if we also introduce the intermediate concept of the r-skeleton Sk(X,r) of a shape X:

(25) 
$$\operatorname{Sk}(X,r) = \bigcap_{\varepsilon > 0} \left( \overline{X}(-r) \left( \overline{X}(-r) \circ \overline{B'}(\varepsilon) \right) \right),$$

then we can rewrite the above property as follows:

(26) 
$$B(x,r)$$
 maximal in  $X \iff x \in Sk(X,r)$ .

Finally, from the union of all r-skeletons we obtain the formula for the central skeleton Sk(X) (Lantuéjoul [208]):

(27) 
$$Sk(X) = \bigcup_{r>0} Sk(X, r) \\ = \bigcup_{r>0} \bigcap_{\varepsilon>0} \left( \overline{X}(-r) - \left( \overline{X}(-r) \circ \overline{B'}(\varepsilon) \right) \right).$$

 $\diamond$  Shape reconstruction

An important property of the skeleton Sk(X) is the ability to *reconstruct* the shape X and all its morphological transforms, if we also know how the radius varies along Sk(X).

• The shape X can be reconstructed from its r-skeletons by dilating them with corresponding disks of radius r:

(28) 
$$X = \bigcup_{r>0} \operatorname{Sk}(X, r) \oplus B(r).$$

• Similarly, the skeleton of the eroded shape X(-r') is made of the *r*-skeletons for r > r':

(29) 
$$\operatorname{Sk}(X(-r')) = \bigcup_{r > r'} \operatorname{Sk}(X, r),$$

so that X(-r') can be reconstructed from the *r*-skeletons with r > r' by dilating them with corresponding disks of radius r - r':

(30) 
$$X(-r') = \bigcup_{r > r'} \operatorname{Sk}(X, r) \oplus B(r - r')$$

• The same holds for dilations, if we define  $X(r) := X \oplus B(r)$ , then:

(31) 
$$X(r') = \bigcup_{r>0} \operatorname{Sk}(X, r) \oplus B(r+r').$$

• For openings, we get:

(32) 
$$X \circ B(r') = \bigcup_{r > r'} \operatorname{Sk}(X, r) \oplus B(r).$$

#### 3. RELATIONS



FIGURE 13. The SKIZ of a configuration of objects (3 black disks in a frame) is the skeleton of its complementary set (here in white). The external part of the SKIZ is induced by the frame.

Under the assumptions of regularity of X stated at the beginning of Section 5.4.2, it can be proved that the skeleton Sk(X) is a closed set of dimension 1, i.e., a graph, composed of regular curve segments that have termination points and are connected at a few independent vertices of finite order. Note that, on the other hand, the skeleton is also an unstable structure that is very sensitive to noise. As we have already emphasized in the previous chapter, small bumps in X's boundary suffice to create extra segments and vertices in Sk(X). This problem can be overcome by multi-scale skeletonization methods.

 $\diamond$  Skeleton by influence zones

The idea of skeleton becomes especially interesting when applied to the complement  $X^c = W - X$  of a configuration of objects  $X = \{X_i\}$  in a domain W. The shapes  $X_i$  are the components of X, and we assume that they are closed, regular, well separated from each other and that  $X^c$  is connected. The skeleton Sk $(X^c)$  is then pruned by removing secondary segments emanating from the singularities of  $X_i$ 's boundaries. This yields the *skeleton by influence zones* or SKIZ (see 2.2.2), i.e., the subset of W containing the points bounding the  $X_i$ 's influence zones (Figure 13).

Formally, the influence zone of  $X_i$  denoted by  $Z(X_i)$  is the set of points that are closer to  $X_i$  than to any other  $X_j$  for  $j \neq i$ :

(33) 
$$Z(X_i) = \{ x \in W \mid d(x, X_i) < d(x, X - X_i) \},\$$

and the SKIZ of X denoted by  $\mathcal{Z}(X)$  is the critical set of boundaries where these zones meet, i.e., the complement of the union of all influence zones:

$$\mathcal{Z}(X) = W - \bigcup_{i} Z(X_i) = W - Z(X) = Z(X)^c.$$

Note that when the components  $X_i$  are points, the SKIZ corresponds to the well-known *Voronoi diagrams*.



FIGURE 14. In this example of an external SKIZ, the triple point to the left codes the fact that there exist locally three parts in interaction. But these three parts are not the three objects: two parts of them are parts of the same non-convex object, while the third object that is 'inside' has no influence on the triple point. The existence of a loop in the right branch of Sk codes the relation of "insideness" and explains the global relational structure of the configuration.

The concept of SKIZ ascribes a morphological content to the gestaltic notion of relations, as exposed at the beginning of this chapter (see Section 1). The SKIZ offers a morphological coding of the background  $X^c$  and, therefore, of the spatial relations between the  $X_i$  components. Relations emerge from the global pattern in the same space as the objects. And such a genesis of relations is valid for complex relations, as it is clear from Figure 14.

Finally, it is easy to see the relation of inclusion between the medial axis skeleton and the skeleton by influence zones:

$$(34) \qquad \qquad \mathcal{Z}(X) \subset \mathrm{Sk}(X^c),$$

since for any  $x \in \mathcal{Z}(X)$  there exists i, j such that x is on the boundary between  $X_i$  and  $X_j$ , which means that  $d(x, X_i) = d(x, X_j) = r$  and therefore that B(x, r) is a maximal disk of  $X^c$ . However, the reverse relation does not hold because the medial axis skeleton Sk $(X^c)$  contains centers of maximal disks that are in contact with only one of the  $X_i$  components.

5.4.3. Implementation. In Chapter 2, Section 9.2, we have seen Hugh Bellemare's implementation of the cut locus algorithm in a neural network. Figure 15 shows the SKIZ of a complex configuration of five objects as an external CL. We present a more sophisticated neural implementation in Section 7.



FIGURE 15. The SKIZ of a complex configuration of 5 objects as an external CL. Bellemare's implementation in a neural network with 5 layers (see Chapter 2, Section 9.2).

# 6. Numerical simulations based on cellular automata

To give a concrete illustration of the above principles and validate our hypotheses empirically we conducted a series of numerical experiments.

Our material consists of simple  $100 \times 100$ -pixel gray-level images, derived from artificial data. They contain various geometrical figures in the style of "contour sketches", e.g., circles, polygons, irregular curves, etc. We assume that they represent a typical outcome of the early stages of visual processing, essentially low-level segmentation and schematization processes. We also assume that each picture is composed of two objects that are already segmented and appropriately tagged by the system: the trajector (TR) and the landmark (LM), respectively acting as figure and background. In most cases, simple contour connectedness is enough to distinguish object domains from each other.

We now describe a perceptual-semantic machine, whose goal is to answer "yes" or "no" given one picture in input and one question about this picture, such as: "Is TR in LM?", or "Is TR above LM?", or "Does TR go across LM?", etc. As said above, the main idea is that a specific question triggers a specific chain of transformation routines, eventually yielding a categorical answer. The boundaries are active since they constitute the main triggering features of the routines. Generally, the TR plays the center role in the sense that it tends



FIGURE 16. Schematic representation of "the ball in the box". TR (the circle) is the ball, LM (the rectangle), the box. We code TR in black and LM in gray. The image frame is a square of size  $100 \times 100$ .

to expand, whereas the LM tends to block this expansion, either as a passive obstacle or by counter-expansion.

# 6.1. Example 1: "the ball in the box"

The first example deals with the "containment" schema. Consider a very simple scene where TR and LM are two closed objects and TR is enclosed inside LM. For example, Figure 16 is a sketch representing "the ball in the box".

6.1.1. Heat diffusion. According to our morphological interpretation of 'in', asking whether "TR is in LM" means checking whether LM blocks TR's diffusion, i.e., whether activity from TR can spread out of LM and reach the boundary  $\partial W$  (the horizon) or the visual frame  $W^6$  in a time  $t \leq t_{max}$  where  $t_{max}$  is the characteristic travel time across W.

A first possibility is to rely on a simple diffusion process such as the heat equation

(35) 
$$-\mu \frac{\partial a}{\partial t} = \Delta a,$$

where a(x, y, t) is the activity rate at point (x, y) at time t, and  $\mu$  a small coefficient. In the absence of specific boundary conditions, this diffusion process relaxes towards a flat landscape  $a = a_0$ , where  $a_0$  is an arbitrary constant. In our case, we clamp the activity on the contours of TR and LM at two different constant values. At any time t, we set:  $a(\partial LM, t) = 0$  and  $a(\partial TR, t) = 1$ , where  $\partial LM$  and  $\partial TR$  represent the set of points (x, y) belonging to LM's and TR's contours, respectively. To simplify, the initial state is also set to 0 everywhere else, although this has no influence on the final equilibrium state. Metaphorically, TR plays the role of an expanding "heat source" (spreading around positive values) and LM a "cold obstacle" (maintaining zero values).

<sup>6</sup> The definition of "visual frame" depends on the context. Generally, it is the visual field but it can also be larger. We will not address this issue here.



FIGURE 17. Heat diffusion process applied to "the ball in the box." (a) Initial state t = 0; (b) iteration t = 100; (c) iteration t = 300; (d) iteration t = 2000. With activity values ranging from 0 to 1, the gray-level code is periodic to make diffusion fronts visible (values 0, .4, .8 coded in white; values .2, .6, 1 coded in black). The contour of TR (the ball) is clamped at 1 and the contour of LM (the box) at 0 (exceptionally coded in gray to keep it visible). The interior of the ball rapidly converges towards 1, the gap between the ball and the box is filled with a gradient (decreasing values from 1 to 0), and the outside of the box remains uniformly 0.

Thus, we postulate that the question "Is TR in LM?" activates the boundaries of TR and LM in a specific way. As a result, the activity landscape relaxes towards a smooth gradient stretching between TR and LM (decreasing from 1 to 0) and, since TR is entirely included in LM, the activity outside of LM remains uniformly 0 (Figure 17).

Thus, the fact that no activity transpires outside of LM's domain and reaches the image boundary  $\partial W$  can be considered as a characteristic categorical clue that TR is 'in' LM. Two counter-examples are shown in Figure 18: the ball is not in the box, whether crossing it or outside of it, and the diffusion process soon reaches the top border of the image frame, where it creates positive activity values.

Again, irrespective of the metric shapes of TR and LM, this type of analytical process reduces the applicability of 'in' to a purely binary (i.e., discrete) categorical criterion: is there or not noticeable activity on the boundary  $\partial W$  before  $t_{max}$ ?

6.1.2. Morphological transformations. As mentioned above, diffusion processes governed by PDEs are rather computationally expensive and not well adapted to practical applications. For example, concretely, it is difficult to estimate how long they take to reach a satisfying state of equilibrium.

We saw in Section 5.4 that more convenient algorithms can be derived from mathematical morphology (MM). The two basic processes of dilation and erosion offer a range of interesting transformations. Shape dilation is simply done by "aggregation": the shape is progressively padded with additional layers



FIGURE 18. Heat diffusion process applied to counter-examples of "the ball in the box". (a) "crossing" example at t = 0 and (b) iteration t = 500; (c) "outside" example at t = 0 and (d) iteration t = 300. Boundary conditions and periodic gray-level code are the same as in the previous figure. In both cases, the activity on the top border of the image frame is rapidly increasing from zero to positive values.



FIGURE 19. Dilation routine applied to "the ball in the box". (a) Initial state t = 0; (b) iteration t = 6; (c) iteration t = 24; (d) stopping at iteration t = 60. Black pixels aggregate onto TR's contour, i.e., wherever at least one of their 4 nearest neighbors is black. LM's contour is clamped on white (although made visible in gray) and eventually stops TR's expansion after a small number of iterations. TR's activity remains entirely confined in LM.

of non-zero pixels, as if each pixel emitted a small disk around itself. Shape erosion does the opposite: non-zero pixels are homogeneously removed from the border, layer after layer. Note that the erosion of a shape is equivalent to the dilation of its complementary. In the simplest version of these routines, images are binary (shape in black, background in white) and a black pixel is added to (or removed from) the shape domain if at least one of its 4 nearest neighbors is black (respectively, white). All pixels are visited once per iteration, so that the number of iterations is roughly equal to the increase (or decrease) in thickness.

Figure 19 shows the dilation process applied to the ball in the box, where LM, the "box", again acts as an obstacle. Similarly to the heat diffusion, the obstacle is created simply by clamping all its pixels on white (although we make it visible in gray for the figures).



FIGURE 20. Cut-locus created by the simultaneous dilation of TR and LM, applied to "the ball in the box". (a) Initial state t = 0; (b) iteration t = 6; (c) iteration t = 11; (d) stopping after iteration t = 21. As in previous figures, TR is coded in black and LM in gray. Black pixels aggregate on TR only if they are not in LM's domain, and conversely. The routine stops when TR stops growing.

Since pixels belonging to the obstacle  $\partial LM$  are not allowed to change to black, the expansion process eventually stops. The number of iterations required to yield an answer is finite and small, making this kind of transformation more practical than physical diffusion. In any case, both algorithms model the same phenomenon—the active expansion of an object against a passive obstacle. Here too, TR is said to be in LM if no activity can be detected on the image frame (categorical binary criterion).

6.1.3. Influence zones and cut-locus. An interesting alternative to the previous scenario is to let LM play an active role, too, and create a counter-expansion toward TR. The resulting figure shows a boundary where the two expansion areas meet (Figure 20). The areas stretching on either side of this boundary are the influence zones of the objects. Here too, the "containment" criterion must distinguish between two types of activity, one coming from TR, the other from LM, and check that TR's activity is not detected on the image boundaries.

Note that TR's contour after expansion (the boundary) constitutes a good approximation of the *external cut-locus* of the scene, which is the generalized symmetry axis of the background (the complementary zone between the ball and the box). This cut-locus (here, looping back on itself) can also be revealed by a skeletonization of the background (Figure 21).

Technically, skeletonization algorithms gradually erode the black domain while preserving local pixel configurations of thickness 1. In a cellular automaton implementation, each one of the 256 possible neighborhoods of  $3 \times 3$  pixels around a black pixel is associated with an update rule: removing (i.e., changing to white) or not the center pixel. Additionally, white pixels always remain white, independently of their neighborhood. On the other hand, the true cutlocus contains all the points that are equidistant from TR and LM (Blum [38]). Figure 22 compares three different methods generating a boundary between T and LM: double dilation, skeletonization, and exact cut-locus.



FIGURE 21. Skeletonization of the background external to TR and LM, applied to "the ball in the box". (a) Initial state t = 0; (b) iteration t = 6; (c) iteration t = 16; (d) stopping after iteration t = 64. Image values are reverted through  $x \mapsto 1-x$  with respect to the previous figure. Note the small internal cut-locus of the ball, reduced to a quasi-point (3-pixel domain).



FIGURE 22. The boundary between TR and LM (shown in black) can be computed in three different ways, applied to "the ball in the box." (a) Approximation obtained by double dilation. (b) Approximation obtained by skeletonization of the complementary area. (c) Exact cut-locus computed by equidistance.

Finally, we will favor the double dilation method as it is faster and less *ad hoc* than the other two. It also naturally corresponds to the expansion phenomena that we are emphasizing. Figure 23 shows an application of this routine to the two counter-examples of "the ball in the box" seen previously (Figure 18).

# 6.2. Example 2: "the bird in the cage"

Consider now a more complicated case, "the bird in the cage", in which the container LM (the cage) has openings. We saw in Section 2 that the English preposition 'in' is neutral with respect to this type of modification. Within the framework of active semantics, such an invariance property is not given but is *constructed* through a preprocessing of LM.

6.2.1. *Preprocessing of LM*. The difficulty here is that the LM's contour is discontinuous. One solution is to perform a preliminary closing of LM before

3. RELATIONS



FIGURE 23. Double dilation applied to two counter-examples of "the ball in the box". TR's activity (in black) reaches the image boundary after 25 and 34 time steps, in (b) and (d) respectively.



FIGURE 24. Detection of the "containment" schema by preliminary closing of a lacunary LM, applied to "the bird in the cage". (a) Initial state t =0; (b) dilation of LM after iteration t = 8; (c) skeletonization of (b) in 18 additional iterations; (d) expansion of TR (here, by diffusion).

expanding TR (whether by diffusion or dilation). The simplest way to treat any arbitrary contour is to compute a dilation followed by a skeletonization, instead of an erosion (Figure 24).

6.2.2. *Global processing by SKIZ.* However, here too, it appears simpler to trigger a double expansion of both TR and LM in parallel, which will reveal the external cut-locus (Figure 25).

Note that, as desired, morphological details of TR have only little influence on the outcome of the dilation. They are erased within a few iterations and, given any original shape, the dilated domain quickly reaches an undifferentiated "blob".

## 6.3. Remarks

To develop a complete morphological analysis of a preposition seemingly as simple as 'in', we would need much finer and more efficient algorithms than the ones just outlined. For example, to take into account situations where LM is not only lacunary but is also not completely surrounding TR, e.g., as in "the fruit in the bowl", a spherical completion  $LM^S$  of LM can be used, or a "solid" expansion of TR that blocks as soon as the diffusion front reaches  $LM^S$ .



FIGURE 25. Detection of the "containment" schema by double dilation, applied to "the bird in the cage." (a) Initial state t = 0; (b) iteration t = 6; (c) iteration t = 14; (d) iteration t = 27.

In even more complex cases, e.g., metonymic cases such as "the spoon in the cup" (where only the spoon head is in the cup, while the spoon handle sticks out; see Section 1.2), higher-level cognitive preprocessing must intervene—here, a stored mereological decomposition of the spoon into two functionally different parts. As the head is considered more central than the handle, 'in' is applied preferentially and metonymically to the head. The reverse situation, in which the handle is inside and the head outside, is a much less good example of "the spoon in the cup."

# 6.4. Example 3: "the lamp above the table"

Let us consider now a different example: the English preposition 'above'.

6.4.1. Preprocessing of LM. The 'above' localizer fundamentally involves an up-down gradient. Its semantics is not only geometric (positional) but, in a qualitative sense, physical. At first, the morphological analysis of the spatial "superiority" schema could thus use an *anisotropic* expansion of TR along the vertical axis, along with detection of activity on the inferior image boundary  $\partial W$  (Figure 26a-b). If the TR finds itself 'above' LM, then LM will block TR's expansion and no activity should reach the bottom (this idea was also developed by Regier [320]). However, to properly handle problematic cases of translation (Figure 26c-d), a preliminary horizontal expansion of LM can be necessary (Figure 26e-g).

6.4.2. Global processing by SKIZ. Here again, the simultaneous and isotropic expansion of both TR and LM participants seems to offer a good solution in most situations. As before, the influence zones and their boundary Z are barely sensitive to variations around the prototypical configuration. In the case of 'above' the categorical binary criterion depends on the intersection of Z with the image boundary  $\partial W$  and the fact that this intersection should not be on the bottom border (Figure 27).



FIGURE 26. Detection of the "superiority" schema by sequential anisotropic expansion, applied to "the lamp above the table". (a)-(b) Prototypical case of TR 'above' LM. (c)-(d) Marginal case involving a problem due to translation. (e)-(g) Solving the translation problem by preliminary horizontal expansion of LM.



FIGURE 27. Detection of the "superiority" schema by double isotropic expansion, applied to "the lamp above the table." (a)-(b) Standard configuration. (c)-(d) Shifted configuration.

Generally, the intersection of Z with  $\partial W$  also provides critical information on the relative location of TR and LM. For example, in the case of 'above', it allows to distinguish among similar cases that different languages conceptualize by different grammatical elements (Figure 28).



FIGURE 28. Different subtypes of "TR above LM", distinguished by the shape of the boundary Z and the location of its two intersection points with the image boundary  $\partial W$ . The two blobs represent TR and LM from top to bottom, respectively. The SKIZ boundary Z was computed exactly by equidistance. (a) Small TR above big LM: Z shows a positive curvature of paraboloid type. (b) TR and LM have similar size: Z is approximately flat. (c) TR is shifted to the side of LM: Z is paraboloid again but tilted so it also intersects a vertical side of the frame. (d) Big TR above small LM: Z has a negative curvature (corresponding to 'par-dessus' in French).

6.4.3. Toward a solution to cognitive linguistic challenges. The previous example clearly shows the possibility of a discrete categorization on the basis of continuously varying spatial relations. They provide a solution to the deep perceptual-semantic problems posed by cognitive linguistics.<sup>7</sup> Another illustration of these principles is given by Figure 29, in which the TR is a disk hovering above a triangle LM. As TR is continuously shifted from left to right, it creates a smooth transition between the concurrent schemata "TR above LM" and "TR beside LM". Once more, the SKIZ boundary Z proves to be useful to extract a categorical criterion from the morphological analysis of this family of scenes. In the 'above' case, Z intersects the image boundary  $\partial W$  only on the top-top border (or, possibly, left-top, top-right, or even left-right), whereas in the 'beside' situation the intersection points are top-bottom (or, possibly, left-left, or right-right).<sup>8</sup> Therefore, the bifurcation from one class to another can be detected when Z reaches (or leaves) the bottom boundary of the image.

#### 6.5. Links with Kosslyn's works

This "above vs. aside" example shows how a morphological approach can solve difficult problems in the relation between perception and language. One of these problems is the difference between *categorical* and *metric* processing

<sup>7</sup> See, e.g., Berkeley's  $L_0$  and NTL (Neural Theory of Language) projects launched in the 1990's by Jerome Feldman and George Lakoff. See [104].

<sup>8</sup> In the case where  $\partial W$  is circular (closer to the real visual field), it can still be divided into top, right, bottom, and left by considering an embedded square. These four directions find natural cues in our universal sense of gravitation and bilaterality.



FIGURE 29. Continuous transition from (a) 'above' to (b) 'beside.' The bifurcation happens when the SKIZ Z touches the bottom boundary of W.

of spatial relations. It has been thoroughly investigated by Stephen Kosslyn [191], who showed that they correspond to hemispheric lateralization. In a special issue of *Neuropsychologia* (44, 2006) dedicated to "New insights in Categorical and Coordinate Processing of Spatial Relations", S. Kosslyn [193] summarizes his theory concerning the "division of labor" of complex cognitive tasks into many simple sub-tasks. The occipital retinotopic areas implement a visual buffer where edge detection and segmentation are processed (see Chapter 2). The ventral "What" pathway goes from the occipital lobe to the inferior temporal cortex and processes the properties of objects. The dorsal "Where" pathway goes from the occipital lobe to the posterior parietal cortex and processes the properties of location. Additionally, a long-term associative memory processes spatial relations between objects and parts of objects. Now, just as the adumbrations of a single object can be infinitely many, the spatial relations in a configuration of objects can vary in a continuous manner, are also infinitely many, and must therefore be *categorized*. Hence, the necessity of a categorical processing of their continuous information.

In a series of papers, David Kemmerer [178] (see also [176], [177]) took into account cross-linguistic studies over more than 6,000 languages and looked for

neuroanatomical correlates of linguistically encoded categorical spatial relations.

He found that such correlates exist and are located in the left supramarginal and angular gyri.

Representations of coordinate spatial relations involve precise metric specifications of distance, orientation, and size; they are useful for the efficient visuomotor control of object-directed actions such as grasping a cup; and they are processed predominantly in the right hemisphere. In contrast, representations of categorical spatial relations involve groupings of locations that are treated as equivalence classes; they serve a variety of perceptual functions, such as registering the rough positions of objects in both egocentric and allocentric frames of reference; and they are processed predominantly in the left hemisphere.



FIGURE 30. Kemmerer's and Damasio's results on the localization of the processing of categorical spatial relations in the left supramarginal gyrus. The letter I (resp. D) indicate blood flow increases (resp. decreases). (From [178], adapted from Damasio et al. [77])

Figure 30, from [178], shows (F = figure and G = ground)

results from a PET study in which English speakers viewed drawings of spatial relations between objects (e.g., a cup on a table) and performed two tasks: naming F, and naming the spatial relation between F and G with an appropriate preposition. When the condition of naming objects was subtracted from that of naming spatial relations, the largest and strongest area of activation was in the left supramarginal gyrus.

Our morphological model can explain how a categorical perception of spatial relations can emerge from continuous variations: the cut locus of the configuration can bifurcate from one criterium ('above') to another ('aside'), and bifurcation is categorical.

# 6.6. Example 4: "zigzagging across the woods"

As our last example, we return to 'across', which was already examined in Section 4. In the case of the "transversality" schema, the morphological analysis has different requirements. As we saw previously, it mostly consists of a tubification of the trajectory of TR (dilation followed by a skeletonization), while extracting the cut-locus of LM, irrespective of the complexity of its texture (Figure 31). The latter can also be carried out by dilation and skeletonization. What remains is to detect the presence of a *quadruple singularity* point, instead of mere activity on the boundaries. This point signals somewhere

#### 3. RELATIONS



FIGURE 31. The "transversality" schema illustrated by "zigzagging across the woods'. (a) The input image; (b)-(c) preprocessing of TR by dilation and skeletonization; (d)-(e) similar preprocessing of LM. Notice in (e) the quadruple intersection point between Sk(TR) and Sk(LM), typical of 'across'.

on the image the crossing between two boundaries of different origins (TR and LM)<sup>9</sup>.

Note the drastic simplification of the data performed by morphological schematization and how the virtual structures they create allow to extract the "transversality" invariant. Morphological algorithms reveal the typical crossing pattern  $\times$ , which can later be easily accessed to produce the desired categorical criterion. Our thesis is that the "TR across LM" *predication* becomes possible only on the basis of this preliminary morphological schematization.

# 7. Wave dynamics in spiking neural networks

## 7.1. 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 recurrent network of coupled spiking units, via

162

<sup>9</sup> This condition is important. Quadruple points of intersection between Sk(TR) and Sk(LM) must be distinguished from mere bifurcation points inside Sk(TR) and Sk(LM) respectively (these are generically triple points).

the expansion processes discussed above (see Figure 33 for a preview). There is a vast cross-disciplinary literature, revived by Winfree [412], on the emergence of ordered patterns in excitable media and coupled oscillators. Traveling or kinematic waves are a frequent phenomenon in non-linear chemical reactions or multicellular structures (see, e.g., Swinney-Krinsky [366]), 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.

#### 7.2. Spatio-temporal 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 linked them with behavior (see, e.g., Abeles [2], Bialek et al. [33], O'Keefe-Recce [247]). *Temporal coding* in the sense introduced by von der Malsburg [397] 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 (Gray et al. [131]).

While most experiments and models involving neural synchronization were based on zero-phase locking among coupled oscillators (e.g., Campbell-Wang [49], König-Schillen [186]), *delayed* correlations have also been observed (Abeles [2], [3]). These nonzero-phase locking modes of activity correspond to reproducible rhythms, or waves, and could be supported by connection structures called *synfire chains* (Abeles [2], Ikegaya et al. [165]). Elie Bienenstock [35] construes synfire chains as the physical basis of elementary "building blocks" that compose complex cognitive representations: synfire patterns exhibit *compositional* properties (Abeles et al. [4], Bienenstock [36]), as two waves propagating on two chains in parallel can lock and merge into one single wave by growing cross-connections between the chains (in a "zipper" fashion). In this theory, spatio-temporal patterns in long synfire chains would thus be analogous to proteins that fold and bind through conformational interactions.

In the present work, we construe wave patterns differently: we look at their emergence on regular 2D lattices of coupled oscillators to implement the expansion dynamics of our morphodynamical spatial transformations. Compared to the traditional blocks of synchronization, i.e., the phase plateaus often used in segmentation models, we are also more interested in traveling waves, i.e., phase *gradients*.

## 7.3. Wave propagation and morphodynamical routines

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 2D 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 (FitzHugh [108]). Each unit *i* is located on a lattice point  $x_i$  and described by a pair of variables  $(u_i, v_i)$ . Unit *i* is locally coupled to neighbor units *j* within a small radius *r* and may also receive an input  $I_i$ :

(36) 
$$\begin{cases} \dot{u}_i = c(u_i - \frac{u_i^3}{3} + v_i + z) + \eta + k \sum_j (u_j - u_i) + I_i \\ \dot{v}_i = \frac{1}{c}(a - u_i - bv_i) + \eta \end{cases}$$

where  $\eta$  is a Gaussian noise,  $d(x_i, x_j) < r$  and  $I_i = 0$  or a constant I. Parameters are tuned as in Figure 32, 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.

Figure 33b shows waves of excitation in a network of coupled BvP units created by the schematic scene "a small blob above a large blob". Block impulses of spikes trigger wave fronts of activity that propagate away from the object contours and collide at the SKIZ boundary between the objects. These fronts are "grassfire" traveling waves, i.e., single-spike bands followed by refraction and reproducing only as long as the input is applied. Under the non-linear dynamics, waves annihilate when they meet instead of adding up. Figure 33a shows the same influence zones obtained by mutual expansion in a cellular automaton, as seen in Section 6.4. Again, there is convincing perceptual and neural evidence for the significant role played by this virtual SKIZ structure and propagation in vision (Kimia [179]).



FIGURE 32. Typical firing modes of a stochastic BvP relaxation oscillator. Plots show u(t) with a = .7, b = .8, c = 3,  $\eta > 0$ , k = 0, I = 0. (a) Sparse stochastic firing at z = -.2 (spikes are upside-down). (b) Quasi-periodic firing at z = -.4. At critical value  $z_c = -.3465$  without noise, there is a bifurcation from a stable fixed point  $u \approx 1$  to a limit cycle.

### 7.4. Two wave categorization models

We now show how wave dynamics can support the categorization of spatial schemata by proposing two models based on the principles discussed previously. In both models, waves implement the expansion-based transformations stated in principles (i)-(ii) (see 2.2.2), then the detection of global activity or of singularities created by the wave collisions is based on principles (iii)-(v). One wave model implements the boundary detection principle (iii) used in the "containment" and "superiority" schemata. The second model focuses on the SKIZ singularities and "signature" detection principle (v), which can be used as a complement or alternative to boundary detection. In this case, we illustrate SKIZ detection with the same 'above' schema as in boundary detection.

7.4.1. Boundary detection with cross-coupled lattices. Detecting the presence or absence of TR activity on the boundaries of the image, as for 'in' or 'above', is not possible in the single lattice of Figure 33b because the waves triggered by TR and LM cannot be distinguished from each other. However, as discussed above and in Chapter 2, a number of models have shown that lattices of coupled oscillators can also carry out segmentation from contiguity by exploiting a simpler form of temporal organization in the lattice: zero-phase synchronization or "temporal tagging". We take here these results as the starting point of our simulation and assume that the original input layer is now split into two distinct sublayers, each holding one component of the scene. Thus, after a preliminary segmentation phase (not presented here), TR and LM are forwarded to layers  $L_{TR}$  and  $L_{LM}$ , where they generate wave fronts separately



FIGURE 33. Realizing morphodynamical routines in a spiking neural network. (a) SKIZ obtained by diffusion in a  $64 \times 64$ , 3-state cellular automaton. (b) Same SKIZ obtained by traveling waves on a  $64 \times 64$  lattice of coupled BvP oscillators in the sparse stochastic regime with  $\eta = 0$ , connectivity radius r = 2.3 and coupling strength k = .04. Activity u is shown in gray levels, brighter for lower values, i.e., spikes u < 0. Starting with uniform resting potentials  $u \approx 1$  (or weak stochastic firing with noise  $\eta > 0$ ), an input image is continuously applied with amplitude I = -.44 in both TR and LM domains. This amounts to shifting z toward a subcritical value  $z = -.3467 < z_c$ , thus throwing the BvP oscillators into a quasi-periodic firing mode. This in turn creates traveling waves in the rest of the network.

(Figure 34). Mutual wave interference and collision is then recreated by crosscoupling the layers: unit *i* in layer  $L_{TR}$  is not only connected to units *j* in  $L_{TR}$  inside a neighborhood of radius *r*, but also to units *j'* in  $L_{LM}$  inside a neighborhood of radius *r'*. The modified dynamics is therefore

$$(37) \ \dot{u_i}^{TR} = F(u_i^{TR}) + k \sum_j (u_j^{TR} - u_i^{TR}) + k' \sum_{j'} (u_{j'}^{LM} - u_i^{TR}) + I_i^{TR}$$

where F(u) is the right hand side of Eq. (36) without k and I terms. A symmetrical relation holds for  $u_i^{LM}$ , swapping TR and LM. Variables  $v_i$  are not coupled and obey equation (36) as before. The net effect is shown in Figure 34: the spiking wave fronts created by TR are canceled by LM's wave fronts and

#### 7. WAVE DYNAMICS IN SPIKING NEURAL NETWORKS



FIGURE 34. Detection of the 'above' schema by mutually inhibiting waves. Two 64 × 64 lattices of BvP units,  $L_{TR}$  and  $L_{LM}$ , are internally coupled and cross-coupled with r = r' = 2.3,  $\eta = 0$ , k = .03, k' = -.03. (a) Single wave fronts obtained by injecting a pulse input I = -.44 in TR and LM for  $0 \le t < 2$ (10 time steps dt = .2). (b) Multiple wave fronts obtained by applying the same input amplitude indefinitely. In both cases, no spike reaches the bottom of  $L_{TR}$ .

never reach the bottom border of  $L_{TR}$ , while hitting the top and partly the sides. This can be easily detected by external cells receiving afferents from the border units and linked to an 'above' response (not shown here). Again, the invisible collision boundary line is the SKIZ, which we now examine more closely in the next network model.

7.4.2. SKIZ signature detection with complex cells. As previously mentioned in Section 6.4.2, detection of boundary activity provides a simple categorization mechanism but is generally not sufficient. Alone, it does not allow to distinguish among similar but non-overlapping proto-schemata, such as the ones in Figure 28. This is where the properties of the SKIZ can help: for example, the concave or convex shape of the SKIZ is able to separate Figure 28a and Figure 28d. As we already emphasized in Chapter 2, Section 9.3, a cut locus is a dynamical structure, and one should also take into account the *flow velocity* along the SKIZ. Indeed, the dynamics of coupled spiking units (Figure 33b) is richer than the morphological model (Figure 33a) 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. The focal point (where the bright band is at its thinnest in

Figure 33b, t' = 32) 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 between a pair of bracketed contours resembling (|) flows inward when facing their concave sides, whereas it flows outward when facing the convex sides of reversed brackets )|(. This refined information can be revealed by wave propagation.

In order to detect the focal points and flow characteristics (speed, direction) of the SKIZ, we propose in this model to introduce additional layers of detector neurons similar to the so-called "complex cells" of the visual system. These cells receive afferent contacts from local fields in the input layer and respond to segments of moving wave bands, with selectivity to their orientation and direction. More precisely, the spiking neural network presented in Figure 35 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. (These are not literally cortical layers but could correspond to functionally distinct cortical areas.) As in the previous model, TR and LM are separated on two independent layers (Figure 35a). In this particular setup, however, there is no cross-coupling between layers and the waves created by TR and LM do not actually interfere or collide. Rather, the regions where wave fronts "coincide vertically" (viewing layers  $L_{TR}$  and  $L_{LM}$  superimposed) are captured by higher feature cells in two direction-selective layers,  $D_{TR}$  and  $D_{LM}$  (Figure 35b), and four coincidence-detection layers,  $C_{1...4}$  (Figure 35c). Layer  $D_{TR}$  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  $D_{LM}$ .

In the intermediate D layers, each point contains a family of cells or "jet" similar to multi-scale Gabor filters. Viewing the traveling waves in layers L as moving bars, each D cell is sensitive to a combination of bar width  $\lambda$ , speed s, orientation  $\theta$  and direction of movement  $\varphi$ . In the simple wave dynamics of the L layers,  $\lambda$  and s are approximately uniform. Therefore, a jet of D cells is in fact single-scale and indexed by one parameter  $\theta = 0, \ldots, 2\pi$ , with the convention that  $\varphi = \theta + \pi/2$ . Typically, 8 cells with orientations  $\theta = k\pi/4$ are sufficient. Each sublayer  $D_{TR}(\theta)$  thus detects a portion of the traveling wave in  $D_{TR}$  (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  $\theta$  (Figure 35b, center of D layers): it sums potentials from afferent L-layer spikes spatially and temporally, and fires itself a spike above some threshold.



FIGURE 35. SKIZ detection in a three-tier spiking neural network architecture. (a) Input layers show single traveling waves as in the previous figure, except that cross-coupling between TR and LM is removed and potentials are thresholded to retain only spikes u < 0. (b) Orientation and direction-selective cells. Each D layer is shown as 8 sublayers (smaller squares) of cells selective to orientation  $\theta = k\pi/4$ , with k = 8...1 from the top, clockwise. (c) Pairwise coincidence cells. Each cell in  $C_1(\theta)$  is connected to a half-disc neighborhood in  $D_{TR}(\theta)$  and its complementary half-disc in  $D_{LM}(\theta - \pi)$ , rotated at angle  $\theta$ . For example: a cell in  $C_1(0)$  (illustrated by the icon in center 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. Same for  $C_2(0)$ , swapping TR vs. LM and upper vs. lower. Similarly,  $C_3(\theta)$  cells are half-connected to  $D_{TR}(\theta)$  and half to  $D_{LM}(\theta - \pi/2)$ , via orthogonal bars (swapping TR/LM for  $C_4$ ). The net output is sparse activity confined to  $C_1(\pi)$ ,  $C_2(0)$ ,  $C_3(3\pi/4)$  and  $C_4(-\pi/4)$ . Note that we use only global  $C_i$  activity: in reality,  $C_1$  cells fire first when the two wave fronts meet, then  $C_2$  cells when they separate again, and finally  $C_3$ and  $C_4$  cells in rapid succession when the arms cross. This precise rhythm of  $C_i$  spikes could also be exploited in a finer model.

Among the four top layers (Figure 35c),  $C_1$  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 halfdisc neighborhood in  $D_{TR}(\theta)$  and the complementary half-disc in  $D_{LM}(\theta - \pi)$ , where the half-disc separation is at angle  $\theta$ . 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 'above' 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  $C_4(-\pi/4)$  (Figure 35c). 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 non-equivalent 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  $C_2(\pi)$ ; 'on top of' yields  $C_2(0)$ like 'above' but no  $C_1$  activity because TR and LM are contiguous (wave fronts can only separate at the contact point, not join); the French preposition 'pardessus' with a convex SKIZ facing up (Figure 28d) yields  $C_3(\pi)$  and  $C_4(-\pi/2)$ , etc. Note that the actual regions of  $C_i(\theta)$  where cells are active (e.g., the location of the SKIZ branches in the south-west and south-east quadrants of Figure 28d) are sensitive to translation and therefore are not good invariant features.

170