

LI LINGUISTIQUE COGNITIVE

VERS UNE SÉMANTIQUE MORPHOLOGIQUE ACTIVE

Jean PETITOT ¹

René DOURSAT ²

1. INTRODUCTION

1.1. Le point de vue cognitif sur la linguistique

Depuis déjà plusieurs années, les nouvelles perspectives dites “cognitives” sur la linguistique ont sensiblement modifié notre compréhension des propriétés syntaxiques et sémantiques des langues naturelles. Les travaux pionniers de Fillmore (1977), Talmy (1978), Jackendoff (1983), Lakoff (1987), Langacker (1987), Sweetser (1990) et autres, ont déplacé les intérêts théoriques pour le langage de ses propriétés syntaxiques formelles et génératives vers ses aspects ressortissant aux *concepts* (ce que l’on appelle la “conceptual structure”) et à la représentation des connaissances. L’“ancrage cognitif” du langage a donc été replacé au centre de l’étude de celui-ci. Il comprend en particulier une prise en considération tout à fait nouvelle (révolutionnaire même) des liens entre le langage, la perception et l’action.

Ces nouvelles orientations linguistiques sont liées à des progrès convergents dans des disciplines connexes. Les neurosciences de la vision et les modèles de vision computationnelle ont déjà fait apparaître différents niveaux de représentation perceptuelle, établissant une distinction entre bas niveaux (précoces ou périphériques) et hauts niveaux (centraux et intégrés). Il est donc devenu pertinent d’analyser techniquement les liens entre la structure spatio-temporelle (topologique, géométrique, morphologique,

¹ C.R.E.A., Ecole Polytechnique, 1 rue Descartes, Paris. e-mail: petitot@poly.polytechnique.fr, <http://www.polytechnique.fr/laboratoires/crea/JeanPetitot/home.html>. Je remercie René Doursat pour l’aide précieuse qu’il m’a apportée en programmant en Java la machine sémantique dont il sera question à la section 6. Sur le plan méthodologique, la synthèse computationnelle est en effet l’une des seules méthodes expérimentales qui soit appropriée à la confirmation / réfutation d’hypothèses cognitives théoriques en linguistique.

² NeuronData. e-mail: rene.doursat@neurondata.com. Cet article a été élaboré lors d’un séjour à l’Ecole Polytechnique. Je remercie le CREA de son soutien et de son accueil.

dynamique) des scènes visuelles et la structure syntactico-sémantique des énoncés qui les décrivent. De même, les progrès dans la compréhension des liens entre perception et action ont conduit à des travaux approfondis sur l’ancrage physique (*embodiment*) des structures conceptuelles. Les conséquences de cette “cognition située” sont fondamentales, en particulier pour la robotique (Brooks, 1986).

Dans cette perspective, le problème de la sémantique des langues naturelles s’est vu complètement transformé. Il ne s’agit plus seulement de comprendre les problèmes de la référence (sémantique logique) ou de l’analyse componentielle lexicale, ou encore des liens avec la syntaxe dans la prédication linguistique, etc. Il s’agit aussi de comprendre *comment le langage s’applique à la réalité perceptive en la catégorisant et en la structurant de façon active*, et comment cette structuration se trouve essentiellement couplée à l’action. On peut par exemple imaginer que, dans un avenir proche, un robot sera capable de déclencher ses programmes moteurs à partir de la description linguistique des images acquises par son système sensoriel et qu’il pourra communiquer cette description à d’autres robots (Steels, 1994).

Or il faut noter qu’il n’existe jusqu’à présent, dans aucune théorie linguistique, aucune explication technique et aucune modélisation mathématique du fait que des mots, même les plus simples, peuvent s’appliquer, non pas au sens dénotatif classique mais en un sens algorithmique effectif, à des contenus perceptifs. Ce problème central a été systématiquement méconnu et ce n’est guère qu’avec certains modèles connexionnistes qu’il a commencé à être abordé en tant que tel. Il faut qui plus est insister sur le fait que, comme l’ont souligné entre autres J. Feldman, G. Lakoff et T. Regier fondateurs du “L0-project” de l’Université de Berkeley qui lui est consacré, il est subtil et *intrinsèquement difficile*.³

Dans cet article, nous nous proposons de traiter, en nous appuyant sur un certain nombre de travaux antérieurs,⁴ certains aspects d’une telle “sémantique perceptive active”. Afin d’éviter de possibles méprises sur le niveau auquel nous nous situons, nous voulons insister d’emblée sur les points suivants :

(i) Nous n’analysons ici que le *noyau morphologique* de certains contenus *prépositionnels*. Mais évidemment l’analyse linguistique de la sémantique prépositionnelle est beaucoup plus complexe. En particulier de nombreuses inférences cognitives top down y opèrent (cf. § 6.4). De même il faut tenir compte du fait que les prépositions peuvent être appliquées à des situations très différentes, grammaticalisent des schèmes universaux et que les invariants qui en constituent les noyaux relèvent d’un

³ Il existe évidemment un grand nombre de modèles de type IA liant langage et perception. Mais la perception y est toujours réduite à des énoncés et à des procédures d’un “langage” de la perception. Elle perd donc son format spatio-temporel. Or c’est précisément celui-ci qui rend le problème difficile.

⁴ Cf. entre autres Petitot [1991b], [1994], [1995].

concept d'espace très abstrait. Bref, il ne s'agit pas d'étudier les structures fines de la prédication linguistique (il existe déjà un grand nombre de théories à ce sujet) mais ses conditions *d'application effective* aux scènes perceptives (conditions qui n'ont été, nous venons de le voir, pratiquement pas explorées). Notre but n'est pas de critiquer les théories linguistiques existantes mais de leur adjoindre une composante *supplémentaire*. Car il faut bien voir que, même à supposer tous les problèmes de sémantique linguistique résolus, on resterait confronté à une difficulté majeure: *dans ses réalisations concrètes les plus typiques, comment un contenu sémantique invariant peut-il être appliqué à des données perceptives infiniment variables?* On ne saurait oublier que la variabilité perceptive possède le cardinal de la puissance du continu!

(ii) La focalisation sur la dimension perceptive ne signifie toutefois en aucun cas que la sémantique des langues naturelles soit de quelque façon que ce soit restreinte à une simple description univoque de la réalité "objective". Même pour les niveaux perceptifs les plus basiques, nous développerons l'idée d'une sémantique *active* qui sélectionne des schèmes structurants et les applique de façon fondamentalement multivoque aux données. Par ailleurs, il est évident, insistons-y, que les schèmes structurants sont d'une grande abstraction et s'appliquent comme tels à des configurations définies dans des *espaces abstraits*. Si nous nous focalisons sur des images concrètes, c'est parce que, déjà à ce niveau (le plus élémentaire qui soit), le problème est tout à fait considérable.

(iii) La prise en compte de la dimension perceptive rencontre immédiatement une difficulté fondamentale, bien connue en reconnaissance des formes, celle de la corrélation entre les aspects symboliques-discrets et analogiques-continus des relations méréologiques liant les parties dans un tout. Qu'il s'agisse de parties d'un objet ou d'objets d'une scène, les constituants d'une totalité organisée sont reliés *à la fois* par des relations catégorielles abstraites et par des relations spatiales (topologiques, morphologiques, métriques) concrètes. Des expériences récentes de psychologie cognitive de la vision (cf. par exemple les travaux de Kosslyn) ont même montré que ces deux types de relations sont traités dans des hémisphères différents. Dans toutes les théories sémantiques classiques c'est le premier aspect (catégoriel) qui est traité. Ici nous nous intéressons aux liens entre les deux aspects en donnant toute son importance au second.

(iv) Une conséquence immédiate de ce point de vue est que la sémantique dont nous allons esquisser certains aspects n'est pas (et ne peut pas être pour des raisons d'essence) linguistique ou lexicale. C'est une sémantique *proto-linguistique*, cognitive, indépendante de toute langue particulière.

(v) Mais bien que proto-linguistique, cette sémantique perceptive doit néanmoins être considérée comme faisant partie intégrante de la sémantique. En effet les théories récentes de la *métaphore* (G. Lakoff, M. Johnson, etc.) et du *blending* (G. Fauconnier, R. Grush, L. Mandelblit, etc.) ont montré que de façon très générale les expressions des

langues naturelles sont construites en faisant opérer des schèmes prépositionnels et actantiels universels sur différents espaces sémantiques. Ce sont de tels universaux relationnels qu'il s'agit d'expliquer à partir de leur enracinement perceptif.

1.2. Le point de vue gestaltiste sur les “relations”

Comme nous venons de le dire, le paradigme classique conçoit les “relations” comme des liens abstraits supplémentaires appliqués à des objets autonomes déjà constitués. Cette conception traditionnelle de type nominaliste se trouve à la base de l'atomisme logique. Le point de vue *gestaltiste*, en revanche, voit les objets et leurs relations comme un tout (unité holistique), duquel on ne peut les extraire qu'ensemble, à la fois les uns et les autres. Une segmentation primaire a pour tâche de distinguer les différents objets, les détachant ainsi de leur fond commun. Ensuite, une analyse méréologique permet de préciser ce “fond”, c'est à dire à la fois les relations que ces objets entretiennent en tant que parties de la scène, et leurs caractéristiques morphologiques. Les relations entre symboles ne sont donc pas données au départ, mais émergent avec les symboles eux-mêmes.

Cette conception a été défendue dès les débuts de la Gestaltthéorie et de la Phénoménologie. Par exemple, dans son ouvrage de 1907 *Ding und Raum* Husserl cherche à comprendre la constitution de “complexes” d'objets. Insistant sur le fait que les sensations visuelles ont la propriété caractéristique de former un *champ* et de posséder un *format* spatial, il explique que tout ce qui possède une unité perceptive se fonde dans l'unité spatiale du champ (et son évolution temporelle). Les objets individués n'existent que détachés et profilés sur leur fond. La conséquence en est que *les relations entre objets se profilent elles-mêmes sur ce fond de l'unité spatiale*. En ce sens, les objets en relation peuvent être considérés comme des parties *d'un objet complexe d'ordre supérieur* qui serait une *configuration*. Les relations méréologiques constitutives des configurations d'objets sont donc *spatio-temporelles*. Et elles doivent être catégorisées en tant que telles.

Un tel point de vue remet drastiquement en cause la “division du travail” traditionnelle entre perception et langage, la perception prenant en charge la forme des objets et le langage leurs relations abstraites. *Les relations sont aussi des formes et, comme les formes, elles peuvent être catégorisées*.

Le point de vue gestaltiste a été particulièrement bien développé par les grammairres cognitives.

1.3. Le point de vue dynamique sur la linguistique cognitive

Si les grammaires cognitives forment une discipline descriptive extrêmement riche, on y constate aussi un certain manque de formalisation et de volonté modélisatrice. Cette lacune est sans doute due en grande partie au contexte de la linguistique théorique de ces dernières décennies. Dominée par le point de vue générativiste formaliste, la linguistique en est venue à assimiler “formalisation” et “langage formel” et, dans la mesure où les nouvelles orientations cognitives rejettent justement l’idée d’un langage formel, la linguistique cognitive a eu tendance à rejeter en bloc le principe d’une formalisation, négligeant l’existence d’autres types de modèles utilisant des outils techniques différents (voir cependant Lakoff, 1988).

Pourtant, des travaux de plus en plus nombreux apparus ces dernières années, en particulier dans le domaine *connexionniste*, mais aussi le courant *morphodynamique* (Thom, 1980; Petitot, 1991a, 1992, 1994, 1995), montrent que des phénomènes cognitifs aussi fondamentaux que la catégorisation, l’apprentissage, l’inférence, l’extraction de règles, etc., peuvent être abordés par des modèles dynamiques physicalistes qui n’ont que peu de points communs avec des modèles de type logico-combinatoire (pour une revue de ces modèles et de la domination actuelle du paradigme dynamiste, voir van Gelder et Port, 1995).

Nous nous proposons de montrer comment une approche dynamique qualitative inspirée (en les généralisant) des propriétés gestaltistes de la perception et de leur possible implémentation neuronale peut être appliquée à une analyse sémantique. Nous partirons des descriptions linguistiques pour en extraire d’abord des procédures dynamiques et parvenir à des algorithmes effectifs, puis pour suggérer que ceux-ci peuvent être naturellement liés à des mécanismes neuronaux impliquant des corrélations temporelles de signaux et des phénomènes de synchronisation (von der Malsburg, 1981; Abeles, 1982; Gray et Singer, 1989; Bienenstock, 1995).

Pour ce faire, nous privilégierons les travaux de Len Talmy, particulièrement intéressants par leur richesse descriptive, leur grande précision de détails et leur systématisme (Talmy, 1978, 1983, 1985, entre autres). Nous nous proposons de reprendre quelques uns de ses exemples comme base pour une modélisation utilisant des routines visuelles de type gestaltiste.

2. LA SÉMANTIQUE “GESTALTISTE” DE LEN TALMY

2.1. Principes généraux

L’une des thèses fondamentales de Talmy est qu’il existe des principes organisateurs de la sémantique linguistique qui sont proches des principes organisateurs

gestaltistes connus en perception et sont, qui plus est, *grammaticalisés*. Autrement dit, les éléments grammaticaux du langage, qui sont en nombre limité et forment des “classes fermées” (par opposition aux éléments lexicaux, nombreux et formant des “classes ouvertes”) seraient responsables de l’organisation conceptuelle sous-jacente au langage. Ron Langacker affirme également que le langage dans son ensemble relève de processus de *catégorisation* qui affectent aussi bien les objets que leurs relations spatio-temporelles (processus, événements). Fondée sur les catégories, la notion de *schématisation* ou de *schème* prend ainsi une importance centrale en linguistique cognitive. En effet, depuis son origine kantienne, *le schème est une procédure permettant de transformer le contenu d’un concept en un algorithme de construction de ses référents*.

2.2. Sémantique active

Une autre hypothèse de Talmy est que, dans un complexe relationnel d’objets où, pour reprendre la terminologie de Langacker, une figure (le *trajector*, *TR*) est profilée sur un fond (le *landmark*, *LM*, ou objet de référence), le fond n’est pas seulement un cadre passif, mais que la géométrie de celui-ci est également conceptualisée grammaticalement. En fait, la majorité des éléments grammaticaux spatiaux traitent la figure comme un simple point, alors que l’objet de référence est conceptualisé de façon plus détaillée et surtout variable selon l’élément grammatical appliqué. Par exemple, le localisateur “dans” de l’énoncé “la balle *dans* la boîte” constitue la boîte en volume, alors que le localisateur “sur” de l’énoncé “la balle *sur* la boîte” la constitue en surface, et celui de “la balle *loin de* la boîte”, en point. Ainsi, différents éléments grammaticaux provoquent différents découpages conceptuels d’une même réalité. Dans les théories classiques, en revanche, la boîte est traitée comme un symbole ne tenant aucun compte de sa forme géométrique et du lien avec la perception, et la sémantique est supposée entièrement encapsulée dans une “signification” des prépositions.

Dans le cadre qui nous intéresse, et pour rendre compte de cette variabilité, nous ferons l’hypothèse que des mots (apparemment) simples comme les prépositions spatiales “dans”, “au-dessus”, “à travers”, etc., sont associés à des *algorithmes de traitement de formes*, qui prennent en entrée les objets perceptifs, leur font subir certaines transformations selon un schème sémantique particulier, et produisent en sortie une réponse, indice de concordance de la scène avec le schème. Nous appellerons ces algorithmes *routines structurantes* ou *routines morphogènes*. On voit à quel point cette approche, qui peut être qualifiée de “sémantique active”, est loin de l’approche lexicale classique de nature définitionnelle et componentielle.

En prenant au sérieux les postulats de ce type de linguistique cognitive, nous parvenons donc à l’idée suivante : supposant données la scène perceptive et la segmentation primaire entre ses principaux acteurs, ou couples forme/fond (résultat d’un

certain nombre d'opérations de prétraitement, essentiellement segmentations et groupements), la sémantique du langage achèvera l'organisation de celle-ci au moyen d'un ensemble de routines structurantes qui

- (i) enrichiront la scène par des structures virtuelles additionnelles (frontières invisibles, points singuliers, lignes de force, anticipation de mouvements, etc.), et, dans le même temps,
- (ii) sélectionneront les traits morphologiques pertinents (surfaces, angles, volumes, segments, etc.) pour ces transformations.

Nous verrons que ces transformations déclenchées par les schèmes sémantiques seront essentiellement de type “*object-centered*”, et utiliseront des routines de *diffusion de contours*.

2.3. Routines structurantes de base

Dans cette section nous donnons quelques exemples de routines structurantes qui se dégagent des exemples de Len Talmy. Il restera ensuite à expliciter l'implémentation algorithmique de telles procédures.

2.3.1. *Invariance d'échelle*

Il existe des propriétés *d'invariance* du langage par rapport à la forme et la structure métrique exacte des données. L'une des hypothèses principales de Talmy est que les éléments grammaticaux sont fondamentalement neutres par rapport à certaines spécifications géométriques, en particulier l'échelle de grandeur. Ceci ne veut pas dire que les éléments grammaticaux sont insensibles à la métrique, mais qu'ils traitent de la même façon des configurations spatiales semblables à des échelles très différentes. Cette invariance par homothétie montre qu'un traitement de type *multi-échelle*, qui existe déjà en vision, doit être inhérent aux routines sémantiques. Les techniques multi-échelles (comme par exemple les ondelettes) sont d'ailleurs devenues un thème central des travaux de vision computationnelle.⁵

2.3.2. *Invariance morphologique*

De nombreux autres exemples montrent que les éléments grammaticaux sont également largement indifférents aux détails morphologiques des objets ou de leurs trajectoires (ces détails pouvant être repris en charge par des éléments lexicaux spécifiques). Ces invariances de sens révèlent que les routines structurantes sous-jacentes

⁵ Cf. Mallat [1989]. Pour les techniques multi-échelles, cf. par exemple Perona-Malik [1990], Whitaker [1993], Morel-Solimini [1995] et Hamy [1997].

vont dans le sens d'une *simplification* des données géométriques. Les routines simplificatrices peuvent être de plusieurs types.

1. Neutralité par rapport aux discontinuités.

Des exemples tels que “la balle dans la boîte”, “le fruit dans la coupe”, et “l’oiseau dans la cage” montrent la neutralité de la préposition “dans” par rapport à l’état de fermeture ou de continuité du contenant. Nous ferons donc les hypothèses suivantes quant aux routines sous-jacentes au concept d’“intérieurité”, ce qui nous amènera à formuler des hypothèses plus générales sur l’ensemble des routines sémantiques.

Dans un premier temps, les objets “figure” (*TR*) et “fond” (*LM*) sont traités séparément. Ici, il s’agira de fermer *LM* en le complétant par des structures virtuelles : la cage est étendue à une surface continue (cf. § 6.3), la coupe complétée par une sphère (cf. § 6.4). Le processus de fermeture consiste donc, en gros, à remplacer les objets par des domaines d’occupation de type “blobs”.

Puis on postule que la routine caractéristique de la relation “dans” consiste à provoquer l’*expansion* de *TR* jusqu’à ce qu’il atteigne l’obstacle constitué par les contours du *LM* fermé. En résumé, une première définition “physique” élémentaire du concept d’intérieurité pourra être : un objet *A* se trouve à l’intérieur d’un objet *B* si et seulement si une expansion isotrope de *A* est stoppée par les frontières de *B*.

Bien que trivial, cet exemple est néanmoins paradigmatique. Il conduit à introduire les règles générales suivantes :

- (i) *un objet possède une tendance spontanée à occuper l’espace tout entier;*
- (ii) *tout objet constitue un obstacle potentiel à l’expansion de ses voisins.*

L’espace partagé commun se trouve alors décomposé en “zones d’influence” des objets, la décomposition pouvant qui plus est dépendre d’une polarisation brisant l’isotropie et orientant les expansions par diffusion. Ainsi, le localisateur “dans” suppose un obstacle isotrope, alors que “au-dessus” (“la lampe est *au-dessus* de la table”) suppose un obstacle situé dans la partie inférieure de la scène, empêchant une expansion verticale de *TR* d’atteindre le bord inférieur du champ visuel.

Donnant un contenu précis à l’idée générale que les objets perceptuels partagent le même espace ambiant, ces règles impliquent que *les relations spatiales entre les objets soient inférées à partir des frontières entre leurs territoires* (fronts de conflit, points singuliers, etc.) *et de l’évolution dynamique de celles-ci.*

2. Neutralité par rapport à la forme.

D’autres exemples de Talmy montrent que des aspects saillants des formes ne sont pas du tout pris en compte par les éléments grammaticaux. Dans la phrase “il a marché à travers la forêt tout droit /en zigzaguant/en courbe”, l’important est que la trajectoire,

qu'elle soit faite en réalité d'une seule droite, de segments ou de courbes, finisse par passer "à travers" la forêt, c'est à dire par joindre deux points (éventuellement virtuels) situés sur deux côtés opposés du domaine boisé. De notre point de vue, ceci montre que la trajectoire subit un processus de schématisation brutal :

- (i) elle est d'abord "convexifiée" ou "tubifiée", par une expansion limitée à un domaine local,
- (ii) puis l'axe principal de ce tube est extrait par "squelétisation", opération équivalente à une expansion inverse (expansion du milieu extérieur), érodant maximalelement l'objet.

Ainsi, du point de vue de la préposition "à travers", la trajectoire originale est considérée comme une fluctuation autour de la direction principale produite par les deux opérations ci-dessus.

La squelétisation est abondamment utilisée en vision computationnelle et implémentée de diverses manières. C'est sans doute le moyen le plus aisé de simplifier une forme tout en conservant ses caractéristiques morphologiques les plus importantes. Nous allons y revenir (cf. sections 4 et 5.4)

3. Neutralité par rapport à l'épaisseur.

Directement reliés aux structures squelétisées se trouvent également des exemples d'invariance du sens selon l'épaisseur des objets, comme dans "la fourmi grimpe le long d'une brindille/d'un tronc/d'une tour". Ainsi, "le long de" est indifférent à toutes les dimensions autres que celle de la direction ou de l'axe principal du *TR*.

3. LE NIVEAU DE LA "TOPOLOGIE COGNITIVE"

Talmy compare souvent ce qu'il appelle la "topologie du langage" (étroitement reliée à la "topologie cognitive" de Lakoff) à la topologie mathématique. En effet, comme on vient de le voir, l'une et l'autre sont neutres par rapport à l'échelle ou les particularités morphologiques. Pourtant, la "topologie linguistique" est à la fois plus abstraite et moins abstraite, ce qui rend difficile sa classification dans la hiérarchie traditionnelle des structures mathématiques. Par exemple, son pouvoir de généralisation est plus grand dans les processus de fermeture (des domaines disjoints peuvent être vus comme une seule surface continue), alors que en revanche certaines proportions métriques doivent être respectées et que les échelles ne peuvent pas être déformées à loisir (pour le langage, un verre est nettement différent d'une assiette, bien que les deux soient homéomorphes). En réalité, la topologie sous-jacente au langage n'a que peu de rapport avec la topologie mathématique.

Rappelons brièvement que les niveaux structurels mathématiques vont du moins contraint au plus contraint, chacun d'entre eux étant associé à une classe particulière

d'applications : au niveau 0, on trouve les structures purement ensemblistes (points indépendants les uns des autres, applications quelconques); au niveau 1, les structures topologiques (points "collés" ensemble par des relations de voisinage qualitatives, applications continues); au niveau 2, les structures différentiables; au niveau 3, les structures conformes (liées aux structures analytiques, applications conformes et holomorphes, préservation des angles); au niveau 4, les structures métriques (distances, isométries) avec métrique locale (espaces riemanniens) ou homogène (espaces euclidiens, hyperboliques, minkowskiens, etc.); enfin, le niveau 5, métrique et linéaire, est celui des espaces vectoriels normés (espaces euclidiens, espaces de Hilbert).

La question serait alors de savoir où s'inscrirait une "topologie cognitivo-linguistique" dans cette hiérarchie. Il semblerait que les niveaux 2 et 3 soient des bons candidats, dans la mesure où ils sont à la fois plus rigides que le niveau topologique 1 et moins contraints que les niveaux métriques 4 et 5. Cependant, il nous semble que cette approche n'est pas la bonne.

En réalité, la conception active de la sémantique conduit à introduire des processus qui s'appuient sur certains niveaux pour monter ou descendre vers d'autres niveaux. Par exemple, l'opération de convexification est une schématisation qui, comme telle, fait perdre beaucoup des propriétés métriques des objets, et les regroupe dans un petit nombre de classes. Une grande diversité d'objets ont des convexifications semblables, et à la limite d'une convexification sphérique, tous les objets deviennent équivalents. Mais dans le même temps, ces classes sont beaucoup plus rigides que celles des objets transformés initiaux, car on ne se permet plus que des transformations qui conservent le caractère convexe des objets. On voit ainsi apparaître un jeu subtil entre appauvrissement et rigidification de structures.

Par définition, être "neutre" par rapport à certaines structures c'est descendre les niveaux dans la hiérarchie ci-dessus. Toutefois, dans l'optique de processus actifs, la neutralité signifie en fait que les objets sont transformés en nouveaux objets *prototypiques* de certaines classes. Et, précisément parce qu'ils sont prototypiques, ces nouveaux objets sont plus rigides. Par exemple, toutes les formes amiboïdes topologico-différentiables (niveaux 1 et 2) deviennent par convexification dynamique (cf. sections 5 et 6) des sphères euclidiennes : en les "sphérisant" pour en faire des prototypes on est donc monté au niveau métrique 4. En résumé, on pourrait dire qu'on a remplacé du "complexe soft" par du "simple rigide". C'est là tout le paradoxe de la "topologie cognitive".

Un autre exemple remarquable de ce phénomène, déjà évoqué, est celui de la squelettisation. Il s'agit d'un processus drastique de simplification mais qui, par ailleurs, est fortement sensible aux propriétés métriques des objets. Un problème intéressant est d'ailleurs de déterminer quelles sont les transformations qui préservent topologiquement le squelette. L'équivalence entre 2 formes : $A \cong B$ ssi $Skel(A) \cong Skel(B)$ est beaucoup

plus faible que l'isométrie. Mais elle est aussi beaucoup plus rigide que les transformations conformes et, a fortiori, que les difféomorphismes et les homéomorphismes.

La sémantique active est donc avant tout affaire de processus, de transformations, de représentations et de recodages. Elle ne correspond pas directement à un niveau de structure mathématique. Elle correspond néanmoins à des opérations qui relèvent toutes d'un même point de vue, que l'on appellera *morphologique*, et c'est en ce sens, et par abus de langage, que l'on peut parler d'un "niveau morphologique".

Thèse : *Ce que l'on appelle "topologie" en sémantique Gestaltiste et en grammaire cognitive correspond en fait à l'action d'opérateurs morphologiques.*

4. OPÉRATIONS SUR LES SCHÉMAS : L'EXEMPLE DE "ACROSS"

Pour montrer la richesse et la difficulté du problème posé par les schèmes percepto-sémantiques, même au niveau de leur noyau invariant le plus élémentaire, nous prendrons, parmi tant d'autres, l'exemple de la préposition "across" (considérée en son contenu proto-linguistique).

4.1. L'invariant de transversalité

Le principal concept géométrique sous-jacent à tous les usages de "across" est celui de *transversalité*. Très intuitif, il a été clarifié par la géométrie différentielle moderne, notamment grâce aux travaux de Hassler Whitney et René Thom.

- (i) intuitivement, deux droites dans le plan \mathbb{R}^2 sont transverses lorsqu'elles sont sécantes (c'est une situation stable, générique);
- (ii) deux courbes dans le plan sont transverses en un point commun si elles n'y sont pas tangentes, i.e. si leurs tangentes en ce point sont des droites transverses;
- (iii) deux droites dans l'espace tridimensionnel \mathbb{R}^3 , même quand elles se coupent, ne sont pas transverses car le fait qu'elles s'intersectent est une situation instable, non générique : il suffit de les déplacer infinitésimalement pour les décroiser;
- (iv) en revanche, une droite et un plan sont transverses s'ils sont sécants (non parallèles);
- (v) une courbe et une surface sont transverses en un point commun si leur droite et leur plan tangents respectifs en ce point sont transverses;
- (vi) deux plans dans \mathbb{R}^3 sont transverses s'ils sont sécants;

(vii) deux surfaces sont transverses en un de leurs points d'intersection, si leurs plans tangents en ce point sont transverses, etc.

Si l'on veut extraire un noyau invariant de ces diverses situations concrètes de transversalité, on voit qu'il correspond à la situation suivante : on dispose d'un espace ambiant M , par exemple \mathbb{R}^2 ou \mathbb{R}^3 , et on considère dans M deux sous-espaces, N_1 et N_2 . Si x est un point commun à N_1 et N_2 , on dira que N_1 et N_2 sont transverses en x si leurs espaces tangents $T_x N_1$ et $T_x N_2$ engendrent tout l'espace tangent en x de l'espace ambiant M , autrement dit si l'on a : $T_x M = T_x N_1 + T_x N_2$.

4.2. Les variantes de la transversalité

Sur la base de ce noyau invariant très abstrait, se développent un certain nombre de schèmes archétypes de la transversalité. D'abord, bien sûr, le "croisement" (au sens de "road-crossing"), mais aussi d'autres schèmes. L'un d'entre eux, que nous allons plus particulièrement étudier, consiste à considérer un domaine muni d'un bord (par exemple une bande) et un chemin qui "traverse" l'objet.

Dans un tel schéma, opère en réalité une double transversalité (figure 1). Il comprend en effet :

- (i) un premier crossing (T_1) à l'entrée du chemin dans le domaine (avec le premier bord),
- (ii) le chemin dans le domaine,
- (iii) une sortie du chemin hors du domaine par un autre crossing (T_2) "opposé" à T_1 .

Ce schème percepto-sémantique correspond à une notion de transversalité entre une courbe 1D et un domaine 2D du plan, reposant elle-même sur une transversalité entre la courbe et le bord du domaine. Son prototype est celui d'une droite transverse à un rectangle.

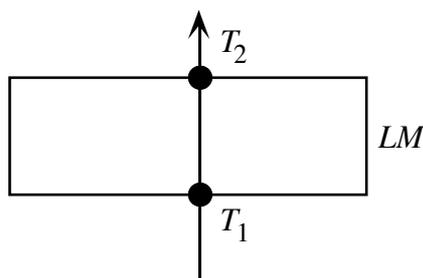


Figure 1: Transversalité d'un chemin 1D et d'un domaine 2D.

4.3. Plasticité du schème

Talmy a donné un vaste ensemble d'exemples d'utilisation de ce schème. Ils sont caractéristiques de la plasticité générale des schèmes percepto-sémantiques.

4.3.1. Plasticité topologique

Il y a d'abord la plasticité topologique : on peut déformer considérablement le chemin T et la forme du contour L sans changer la propriété de "across" à condition de préserver les 3 critères énoncés plus haut (figure 2).

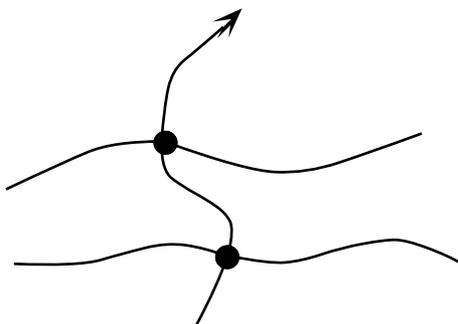


Figure 2: Plasticité topologique de la transversalité.

Mais cette condition nécessaire n'est pas suffisante. On peut facilement mettre à jour d'autres contraintes montrant qu'il n'est pas possible de déformer le domaine D de n'importe quelle manière. En fait, il est assez difficile de déterminer ces limites. Nous allons y revenir. Par exemple la figure 3 ne correspond pas à "across" parce que T_2 n'est pas sur le bord "opposé" à T_1 . Mais qu'est-ce que, qualitativement, un "bord opposé" ?

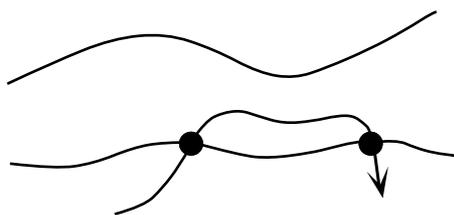


Figure 3: Limites de la transversalité.

4.3.2. La dialectique concrétisation/virtualisation

De fort nombreux exemples montrent que le schème n'a pas besoin d'être complet et que plusieurs de ses éléments peuvent être virtuels, manquants, ou fictifs. Ils forment un éventail entre deux cas extrêmes :

- (i) celui où les bords, donc T_1 et T_2 , sont virtuels (comme dans “the tumble weeds are swept across the prairie”),
- (ii) celui où le chemin est virtuel (comme dans “the snail crawls across the car”).

4.3.3. *L’ambiguïté 2D/3D*

Talmy donne aussi des exemples *d’application adaptative* des schémas. On peut dire “across / over the plateau” mais aussi “across / through the wheat field”, ce qui semble indiquer que l’application de “across” exige que le chemin soit surfacique. Il est assez énigmatique au prime abord que de telles finesses dimensionnelles puissent faire partie de la sémantique des langues naturelles. Mais en fait cela se comprend facilement à partir de la condition de transversalité. Quand le domaine F est $2D$ (“the plateau”), la condition de transversalité porte sur le *bord* de F et fonctionne donc entre courbes. Si F s’épaissit et devient $3D$ (“the wheat field”), c’est alors la *surface* de F qui se met à fonctionner comme bord de son volume. La transversalité doit alors être une transversalité $3D$ entre une courbe et une surface.

On voit bien sur ces exemples simples que la sémantique morphologique active n’est pas “objective” (au sens standard du terme) mais “sélective”: elle *sélectionne* des schèmes morphologiques qu’elle adapte aux données perceptives. C’est en ce sens qu’elle est *active*. On voit également *qu’il existe une compétence géométrique fine sous-jacente à la compétence sémantique*.

4.3.4. *Contraintes sur les déformations*

Nous avons vu plus haut qu’il existait des contraintes additionnelles et que la plasticité du schème n’était pas uniquement topologique. Pour mesurer la subtilité du problème, prenons l’exemple de “traverser un lac”.

- (i) Un chemin qui pénètre un peu dans le lac, en suit le bord et ressort un peu plus loin ne correspond pas à une situation véritablement transversale, mais plutôt à une situation approximativement tangente (où “approximativement” renvoie à la notion d’échelle). Un peu comme si le bord possédait une certaine épaisseur et était identifié à l’un de ses voisinages tubulaires dont l’échelle serait déterminée par le contexte (figure 4).

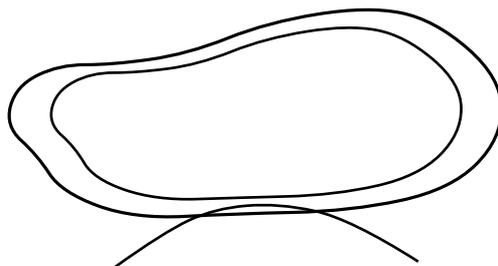


Figure 4: Fausse transversalité assimilée à une tangence approximative.

(ii) Un chemin qui entre dans le lac, arrive presque à l'autre bord et fait demi-tour pose un problème plus subtil. On peut faire l'hypothèse que si "across" n'est pas adapté c'est parce que le second crossing T_2 n'est pas de "l'autre côté" du contour relativement à T_1 . Mais, encore une fois, comment définir "l'autre côté", le "bord opposé"? (figure 5).

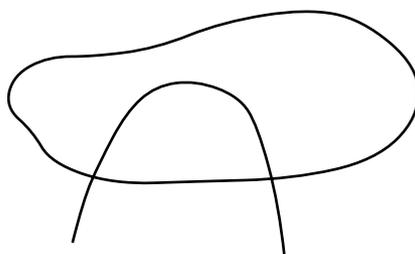


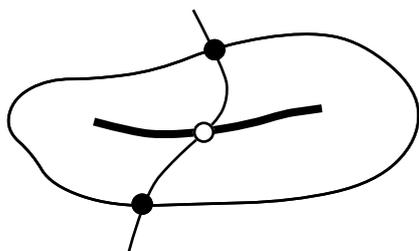
Figure 5: Bien que la situation soit topologiquement équivalente à celle de la figure 1, il n'y a pas transversalité au sens de "across". Pourquoi?

On rencontre là un exemple typique des difficultés soulevées par la nature morphologique de la perception. On ne peut pas les résoudre au moyen de matchings métriques. La méthode serait beaucoup trop coûteuse en raison de la prolifération des templates nécessaires (puisque l'on ne connaît pas la forme en jeu). Quant à un matching moins contraint, il identifierait des situations que l'on ne veut pas identifier (comme les figures 1 et 5).

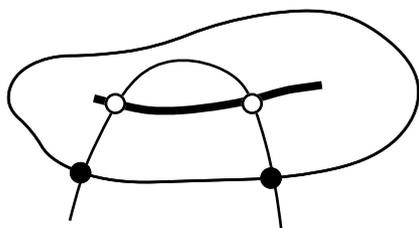
La solution consiste à associer à la situation *des éléments virtuels qui gardent la mémoire du métrique sans être eux-mêmes de nature métrique*. La différence qualitative entre "ce côté-ci" et "l'autre côté" d'un lac, qui est prise en charge linguistiquement par "here" and "there", correspond géométriquement à *l'introduction d'un bord virtuel intermédiaire* situé grosso modo au milieu des deux rives. Avant ce bord, on se trouve "de ce côté-ci" et après ce bord on se trouve "de l'autre côté". C'est ce bord virtuel qui établit le lien entre l'opposition linguistique "here/there" et la situation physique-perceptive.

On remarque alors que le bord virtuel possède bien un aspect métrique, puisqu'il est "au milieu", mais que, en même temps, son rôle est purement qualitatif puisqu'il consiste seulement à partager l'espace en deux parties. Autrement dit sa *position* est quantitativement précise, mais sa *fonction* n'est que de discrétiser le continu sous-jacent. Il permet par conséquent de résoudre (du moins dans ce cas) le problème du passage du continu métrique au catégoriel discret.

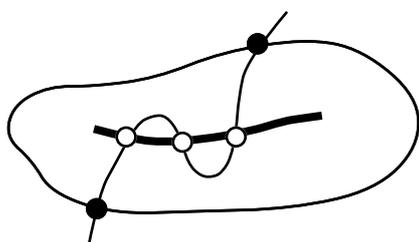
Nous retrouvons ainsi tout naturellement l'idée fondamentale de *cut-locus*, ou de squelette, c'est à dire d'un axe de symétrie généralisé. Sa prise en considération comme structure organisatrice virtuelle permet de résoudre facilement le problème posé (du moins, partiellement). On peut poser en effet que "across" est employé correctement lorsque le chemin relie 2 points du bord qui se trouvent *de part et d'autre du cut-locus* et qu'il est en outre *transversal* à ce dernier. *L'invariant de la transversalité pure se retrouve donc au niveau de l'intersection du chemin et de l'axe de symétrie.* Son archétype continue à opérer, mais sur des *formes transformées*. D'où la plasticité du schème : le chemin peut zigzaguer autour du cut locus, l'important est que le bilan du parcours soit une traversée (figure 6). Nous voyons ainsi émerger l'idée que, pour atteindre les contraintes dont on a besoin, une bonne méthode consiste à rajouter des structures virtuelles supplémentaires.



(a) Il y a bien "traversée" car les 2 points A et B d'intersection transversale du chemin γ et du bord β (points noirs) sont situés de part et d'autre du cut-locus C .



(b) Il n'y a pas traversée car A et B ne sont pas situés de part et d'autre de C et γ intersecte C transversalement en 2 points (points blancs) dont le bilan algébrique $+1-1=0$ n'est pas celui d'une traversée.



(c) Il y a traversée car A et B sont situés de part et d'autre de C et γ intersecte C transversalement en 3 points (points blancs) dont le bilan algébrique $+1-1+1=+1$ est celui d'une traversée.

Figure 6: Résolution du problème de la figure 5.

4.4. L'exemple de l'ambiguïté Across/Along

Un autre exemple de l'intérêt qu'il y a à utiliser un analyseur morphologique de type cut locus pour résoudre certaines subtilités sémantiques est fourni par la différence entre "swimming across the swimming pool" et "swimming along the swimming pool" (figure 7).

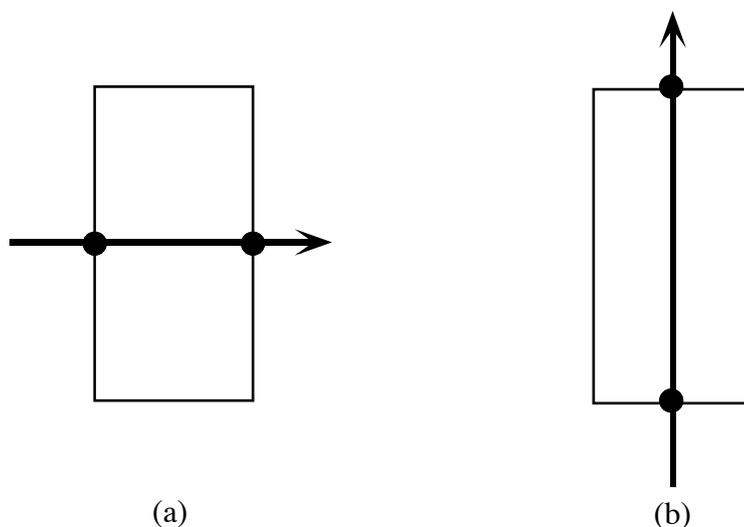


Figure 7: Pourquoi dans le second cas c'est la proposition "along" qui doit être appliquée aux données et non pas la préposition "across"? Les deux situations sont pourtant qualitativement équivalentes.

Dans les deux cas les conditions d'application de "across" sont satisfaites. Pourtant c'est "along" qui sera normalement utilisé dans le second cas. Pourquoi? Bien sûr, on peut toujours introduire un critère métrique supplémentaire et ad hoc et, par exemple, discriminer entre (a) et (b) en corrélant le rapport des longueurs des côtés avec la direction du mouvement. Mais cela n'est pas satisfaisant car on perd le lien avec l'invariant de transversalité et avec le problème d'une catégorisation binaire.

Mais si l'on introduit le cut-locus C du landmark, alors la discrimination entre les deux cas (a) et (b) devient immédiate. Dans le cas (a) le chemin est transversal à C alors que dans le cas (b) il est parallèle à C (figure 8).

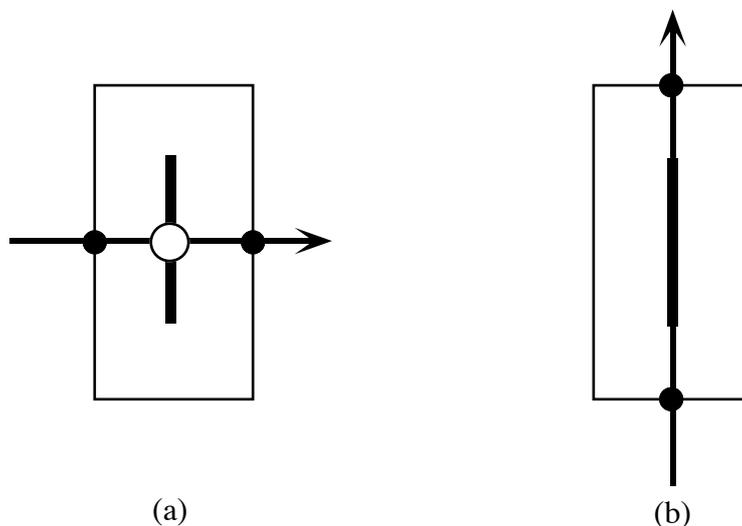


Figure 8: Explication de l'opposition entre "across" et "along".

Évidemment, ce critère rencontre des difficultés dans les cas limites, par exemple dans le cas (a) lorsque le chemin est trop près des bords horizontaux et ne peut plus intersecter *C*. Mais il est néanmoins plus efficace et plus pertinents que les autres.

4.5. D'autres exemples de structures virtuelles : "fictive motions"

L'introduction d'éléments structurants virtuels, fictifs et subjectifs est cruciale pour toute la Gestalt semantics de Talmy. Celui-ci a donné tout un ensemble d'autres exemples de structures fictives prises en charge grammaticalement par le langage, en particulier des exemples de mouvements fictifs que nous rappellerons brièvement afin de montrer la richesse du thème.

- (i) *Emanation paths*. Lorsque l'on pointe vers un objet, l'index construit une ligne virtuelle dans l'espace le reliant à l'objet.
- (ii) *Shadow paths*. Une ombre se "projette" sur un mur comme si elle parcourait un chemin entre l'objet et le mur.
- (iii) *Virtual processes*. En général, une situation statique est conçue comme le résultat d'un processus dynamique virtuel: "the palm trees clustered around the oasis", "the mountains are scattered", etc.
- (iv) *Virtual fronts*. Tout objet définit un "front" constitué d'un plan virtuel et d'un chemin d'émanation perpendiculaire à ce plan.

5. MODÉLISATION: PRINCIPES ET ALGORITHMES

5.1. Computation gestaltiste

L'ensemble des problèmes que nous venons d'évoquer sont des problèmes de Gestalt au sens fort et technique du terme. Ils mettent en jeu une structure globale et des éléments organisateurs virtuels, une dialectique du quantitatif et du qualitatif, du métrique et du catégoriel, ainsi que des phénomènes de plasticité topologico-morphologique. Ils amplifient les phénomènes de la Gestalt classique (comme les contours virtuels ou la *Prägnanz* des bonnes formes) et, sur le plan algorithmique, soulèvent un dilemme ardu de type local/global : trouver l'équilibre entre Gestalt globale et computations locales.

Le problème est en effet le suivant : quelles procédures computationnelles sont-elles en mesure de produire (par une sorte de "scanning") de telles structures organisatrices virtuelles? La difficulté réside essentiellement dans le fait que les éléments globaux (holistiques) et informationnellement infinis (continus) doivent être traités par des algorithmes locaux, n'utilisant que des procédures informationnellement finies. Les grammaires cognitives, de ce fait, sont loin d'être satisfaisantes : elles s'appuient sur les propriétés holistiques mystérieuses de la vision, sans chercher à les expliciter, les modéliser et en extraire des algorithmes.

Qui plus est, les procédures doivent dériver de principes suffisamment généraux pour pouvoir expliquer *l'énigme de la possibilité d'un apprentissage non supervisé* des remarquables subtilités percepto-sémantiques de classes fermées comme les systèmes prépositionnels.

5.2. Opérations structurantes et expansions

Or les opérations structurantes que nous avons esquissées dans la section 4 pointent toutes vers un type bien précis de processus : si l'on veut construire des structures globales en n'utilisant que des mécanismes locaux, la seule solution semble être, d'une façon ou d'une autre, de *déclencher des processus de diffusion-propagation*, ou *expansions*, à partir des données initiales saillantes, en particulier les contours. Ceux-ci deviennent actifs (excitateurs ou inhibiteurs), et déclenchent des "ondes" d'activité. Ces ondes engendrent alors elles-mêmes des singularités de fronts d'ondes et, s'il y a plusieurs sources, des conflits entre fronts.

Récapitulant les routines rencontrées dans l'exploration des exemples précédents (convexification, tubification, squelettisation, cut-locus, bords déclencheurs et bords obstacles, etc.), les deux idées forces que nous proposons pour la sémantique morphologique active sont les suivantes :

- les contours sont des éléments actifs qui, soit déclenchent des diffusions-propagations d'activité, soit inhibent celles en provenance d'autres sources;
- les processus de diffusion-propagation engendrent à leur tour de nouvelles entités géométriques, en particulier des singularités, qui fonctionnent comme des éléments structurants virtuels.

5.3. Algorithmes morphologiques

Les hypothèses précédentes sur une sémantique gestaltiste doivent pouvoir être testées empiriquement au moyen d'une synthèse computationnelle. Mais celle-ci se heurte immédiatement à une difficulté pratique. En effet, les algorithmes standard de diffusion-propagation utilisent des équations aux dérivées partielles qui demandent beaucoup de puissance de calcul et ne sont donc pas pratiquement efficaces. Pour être opératoire, il faut par conséquent en utiliser des versions simplifiées computationnellement plus efficaces.

De tels algorithmes se rencontrent dans ce que l'on appelle la *morphologie mathématique* développée dans les années 70 à l'École des Mines de Fontainebleau par G. Matheron et de son école (J. Serra, M. Schmitt, J. Mattioli, etc.).⁶

5.3.1. Morphologie mathématique des images binaires

L'idée directrice est d'analyser une image au moyen d'*éléments structurants* en utilisant des opérations ensemblistes booléennes. Soit X une image binaire ⁷ 2D dans une fenêtre W (centrée sur l'origine des coordonnées). X est définie par une fonction caractéristique φ sur W à valeurs dans $\{0,1\}$ et correspond à la partie de W où $\varphi(x)=1$. $\varphi^{-1}(0)$ est un fond sur lequel les composantes connexes de X se détachent comme des objets.

En général, l'élément structurant de base est un *disque* B (de rayon r centré sur 0). B doit être considéré non seulement comme un élément *géométrique* (métrique) mais aussi comme un élément *dynamique* résultant d'une diffusion du centre 0 pendant un temps caractéristique. La différence fondamentale avec une EDP de diffusion est que l'élément demeure *binaire*, sans lissage de son bord. On parle donc de *diffusion binaire*. C'est une version très simplifiée de la diffusion réelle, qui est très bien adaptée à l'analyse morphologique.

⁶ Cf. Serra [1982]. Dans cette section nous suivrons essentiellement Schmitt-Mattioli [1994]. Je remercie Jean-Michel Morel de m'avoir fait connaître ces algorithmes. Mon approche morphologique étant liée aux travaux de René Thom, elle reposait sur des modèles de géométrie différentielle et d'EDP difficiles à implémenter de façon efficace (JP).

⁷ Binaire signifie qu'il n'y a que 2 niveaux de gris: blanc et noir (ou $\{0,1\}$).

Il est alors très facile de définir la *dilatation* de X par B , notée $X \oplus \overset{\vee}{B}$. Soit B_u le translaté de B par u : $B_u = \{b + u \mid b \in B\}$. On a bien sûr $X \cap B_u \neq \emptyset \Leftrightarrow$ il existe $x \in X$ et $b \in B$ tels que $x = b + u$ i.e. $u = x - b$. Soit $\overset{\vee}{B} = \{-b \mid b \in B\}$, (si B est symétrique par rapport à l'origine 0 alors $\overset{\vee}{B} = B$). En considérant les points de X comme des sources de diffusion binaire, on obtient la définition naturelle de la dilatation:

Définition.

$$\begin{aligned} X \oplus \overset{\vee}{B} &= \text{ensemble des } u \text{ tels que } B_u \text{ intersecte } X \\ &= \{u \mid B_u \cap X \neq \emptyset\} \\ &= \{x - b \mid x \in X, b \in B\} \\ &= \bigcup_{b \in \overset{\vee}{B}} X_b = \bigcup_{x \in X} \overset{\vee}{B}_x. \end{aligned}$$

La dilatation d'une image X par le disque B la régularise. Elle connecte les objets qui sont suffisamment proches, obture les petits trous (petits relativement à la taille de B) et dilate la taille de X de r (le rayon de B). Bien sûr, les résultats dépendent de la taille de B et des différentes manières d'implémenter l'algorithme sur des grilles digitales (hexagonales, carrées avec 4 voisins, carrées avec 8 voisins, etc.). La figure 9 donne un exemple (adapté de l'icône de Java).



Figure 9: Un exemple de dilatation (grille carrée, #voisins=2, #itérations=1).

L'opération de dilatation possède de nombreuses bonnes propriétés. Par exemple elle est distributive par rapport à l'union: $X \oplus (B \cup B') = (X \oplus B) \cup (X \oplus B')$ et sous-distributive par rapport l'intersection: $X \oplus (B \cap B') \subset (X \oplus B) \cap (X \oplus B')$. Elle est aussi associative $(X \oplus B) \oplus B' = X \oplus (B \oplus B')$ ⁸

Une fois qu'on dispose de cette routine de base, on peut facilement définir d'autres routines.

(i) **Érosion.** L'érosion de X , $X \ominus \overset{\vee}{B}$, est l'opération duale de la dilatation. Elle correspond à la dilatation du complémentaire de X , $X^c = W - X$:

⁸ Pour des précisions, cf. Serra [1982].

$$X \ominus \overset{\vee}{B} = \left(X^c \oplus \overset{\vee}{B} \right)^c = \left\{ u \mid X^c \cap B_u \neq \emptyset \right\}^c = \left\{ u \mid X^c \cap B_u = \emptyset \right\} = \left\{ u \mid B_u \subset X \right\} = \bigcap_{b \in \overset{\vee}{B}} X_b$$

L'érosion d'une image X par un disque B possède les propriétés complémentaires de celles de la dilatation. Elle disconnecte les étranglements assez étroits, efface les objets trop petits et contracte la taille de X de r (voir figure 10).



Figure 10: Un exemple d'érosion (grille carrée, #voisins=4, #itérations=2).

(ii) **Ouverture.** L'ouverture de X , notée X_B , consiste en une érosion suivie d'une dilatation (figure 11):

$$X_B = \left(X \ominus \overset{\vee}{B} \right) \oplus B.$$

Comme l'érosion, l'ouverture efface les objets trop petits et disconnecte les étranglements suffisamment étroits. Mais après avoir érodé elle dilate, ce qui fait qu'elle régularise les objets et renormalise leur taille.

(iii) **Fermeture.** La fermeture X^B d'une image X est l'opération réciproque de l'ouverture. Elle consiste en une dilatation suivie d'une érosion (figure 11):

$$X^B = \left(X \oplus \overset{\vee}{B} \right) \ominus B.$$



Figure 11: Exemples d'ouverture et de fermeture.

La fermeture est exactement l'opération morphologique dont on a besoin pour comprendre comment des bords discontinus peuvent être traités comme des bords continus (figure 12).

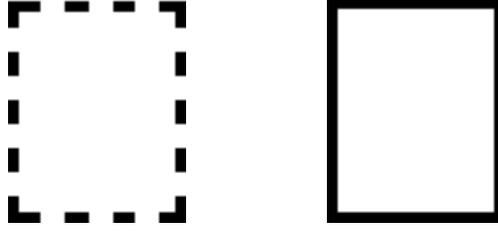


Figure 12: Fermeture d'un bord discontinu: dilatation puis érosion.

En itérant et alternant de telles routines élémentaires on peut analyser une image de plusieurs façons différentes. Nous en verrons à la section 6 quelques exemples.

5.4. Squelétisation

5.4.1. Squelettes et reconstruction de formes

En utilisant les routines morphologiques, il est aisé de calculer le *squelette* $Sk(X)$ d'une forme X (qui est très difficile à calculer en utilisant des EDPs de propagation). Par définition, $Sk(X)$ est le lieu des centres des disques maximaux inclus dans X . Afin d'éviter des pathologies topologiques on suppose que X est ouvert, connexe, avec un bord ∂X différentiable par arcs et une fermeture \bar{X} compacte.⁹ Soit $B(x,r)$ (resp. $B^\circ(x,r)$) le disque fermé (resp. ouvert) de centre x et de rayon r (on note $B(r) = B(0,r)$ et $B^\circ(r) = B^\circ(0,r)$). Soient $X_r = X \ominus B(r)$ et $\bar{X}_r = \bar{X} \ominus B^\circ(r)$ les images érodées correspondantes. $B^\circ(x,r)$ est maximal dans X si l'inclusion $B^\circ(x,r) \subset B^\circ(x',r') \subset X$ implique l'égalité, c'est-à-dire $x = x'$ et $r = r'$. Or il est facile de montrer que le fait que $B(x,r)$ soit maximal est équivalent au fait que x appartienne à \bar{X}_r sans appartenir à aucune ouverture de \bar{X}_r par un disque fermé, i.e. équivalent à $x \in \bigcap_{\varepsilon > 0} (\bar{X}_r \setminus (\bar{X}_r)_{B(\varepsilon)})$. On obtient ainsi la formule de Lantuéjoul (1977) permettant de construire $Sk(X)$:

$$Sk(X) = \bigcup_{r > 0} \bigcap_{\varepsilon > 0} (\bar{X}_r \setminus (\bar{X}_r)_{B(\varepsilon)}) .$$

Soit $Sk_r(X)$ le r -squelette lieu des centres des disques maximaux de rayon r . On a :

$$Sk(X) = \bigcup_{r > 0} Sk_r(X) .$$

⁹ Si $X \subseteq M$ est un sous-ensemble d'un espace topologique M , on note \bar{X} sa fermeture (le plus petit fermé contenant X) et X° son intérieur (le plus grand ouvert contenu dans X).

Cette formule montre que la connaissance du squelette Sk et de la fonction rayon le long de Sk permet de *reconstruire* la forme X , ses dilatations, ses érosions et ses ouvertures. En effet :

(i) X peut se reconstruire à partir des r -squelettes en les dilatant d'un disque de rayon

$$r : X = \bigcup_{r>0} Sk_r(X) \oplus \overset{\circ}{B}(r).$$

(ii) Le squelette de l'érodé X_a est composé des r -squelettes pour $r>a$, i.e. est donné par $Sk(X_a) = \bigcup_{r>a} Sk_r(X)$ et on reconstitue donc X_a en dilatant les r -squelettes pour $r>a$

d'un disque de rayon $r-a$:

$$X_a = \bigcup_{r>a} Sk_r(X) \oplus \overset{\circ}{B}(r-a).$$

(iii) Pour les dilatations, on a de même :

$$X \oplus \overset{\circ}{B}(a) = \bigcup_{r>0} Sk_r(X) \oplus \overset{\circ}{B}(r+a)$$

(r -squelettes dilatés de disques de rayon $r+a$) et pour les ouvertures

$$X_{B(a)} = \bigcup_{r>a} Sk_r(X) \oplus \overset{\circ}{B}(r).$$

Sous les hypothèses de régularité de X , on peut montrer que le squelette $Sk(X)$ est un fermé de dimension 1 — un graphe — composé de lignes régulières possédant des points de terminaison et se connectant à travers des vertex isolés d'ordre fini.

Remarque. Il est bien connu que le squelette est une construction instable très sensible au bruit. On peut surmonter cette difficulté en considérant des squelettes *multi-échelles*.

5.4.2. Squelettes par zones d'influence (*Skiz*)

L'idée de squelette devient particulièrement intéressante lorsqu'on l'applique au complémentaire $X^c=W-X$ d'une configuration d'objets $X=(X_i)$ dans un domaine W (les X_i sont les composantes de X et on les suppose fermées, régulières, bien séparées les unes des autres; on suppose également que X^c est connexe). On simplifie le squelette en enlevant les segments de $Sk(X^c)$ qui émanent des points singuliers des bords des composantes X_i et en ne gardant que les segments qui séparent les extensions respectives des X_i . Cela conduit à la construction suivante, due à Prewitt (1966).

La *zone d'influence* de chacun des objets X_i est définie comme le domaine de W constitué des points x dont la distance à X_i est inférieure à la distance aux autres objets X_j , $j \neq i$:

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

Le squelette par zones d'influence $Skiz(X)$ est alors l'ensemble critique associés aux zones, i.e. le lieu de la séparatrice de conflit séparant les zones d'influence (figure 13) :

$$Skiz(X) = W - Z(X) \text{ avec } Z(X) = \bigcup_i Z(X_i).$$

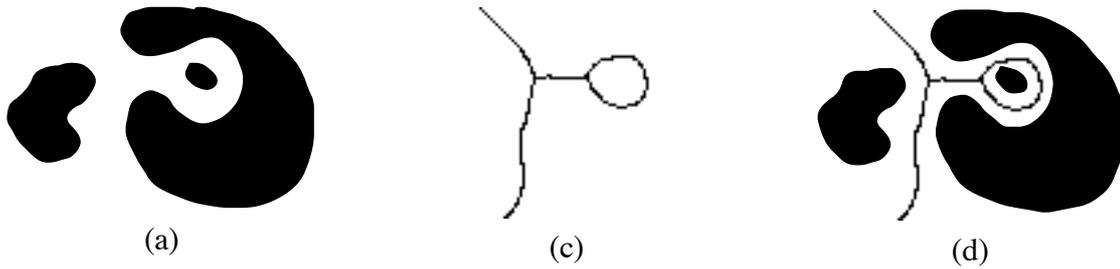


Figure 13: Skiz d'une configuration. (a) la configuration; (b) le Skiz; (c) le partage de l'espace commun en zones d'influence par le Skiz.

Les *Skiz* permettent de conférer un contenu morphologique au point de vue *gestaltiste* sur les “relations” expliqué à la section 1.2. En effet ils codent morphologiquement le background de la configuration X , et donc les relations spatiales entre les composantes X_i . Lorsque les X_i sont des points on obtient les célèbres diagrammes de Voronoi.

Pour $x \in Skiz(X)$ (par exemple x appartenant à la séparatrice entre $Z(X_i)$ et $Z(X_j)$) on a $r = d(x, X_i) = d(x, X_j)$ et $B(x, r)$ est un disque maximal de X^c . On a donc $Skiz(X) \subset Sk(X)$. Mais en général l'inclusion est stricte. En effet, il existe en général dans $Sk(X)$ des centres de disques maximaux n'ayant de contact qu'avec un seul X_i .

6. EXEMPLES ET SIMULATIONS NUMÉRIQUES

6.1. Une machine percepto-sémantique virtuelle

Pour donner une illustration concrète des principes ci-dessus et tester empiriquement nos hypothèses, nous avons conduit une série d'expériences numériques. Notre matériel de travail est un ensemble de petites images d'une taille de 100×100 pixels en 256 niveaux de gris. Les données sont des figures artificielles que nous supposons à peu près représentatives du résultat des premières étapes du traitement visuel, essentiellement une extraction de contours (le “2-1/2D sketch” de Marr, 1982). On fait aussi l'hypothèse que ces sketches sont associés à deux entités, le *trajector* (TR) et le *landmark* (LM), déjà individué, séparés du fond et distingués l'un de l'autre par un processus de segmentation puis étiquetés comme tels. Rappelons que, dans la plupart des cas, la simple connexité des segments de contour peut suffire à séparer deux objets. De nombreux modèles de vision computationnelle, comme des équations de diffusion non-linéaire (Morel et Solimini, 1995; Whitaker, 1993) ou des modèles variationnels

(Mumford, 1993), etc., mais aussi des modèles connexionnistes, essentiellement à base de réseaux d'oscillateurs couplés (voir par exemple König et Schillen, 1991), ont démontré la possibilité d'une segmentation a priori de type "bottom-up" et "data driven". Sur ce plan, notre modèle est donc tout à fait standard: il suppose qu'il existe *une préparation perceptive de l'input* (un peu comme en physique les protocoles expérimentaux permettent de préparer le système étudié et de le placer dans un certain état initial).

Nous décrivons à présent l'architecture et le comportement une "machine percepto-sémantique" dont le but serait de répondre par "oui" ou par "non" à une question posée sur une image donnée en entrée. Par exemple : "Is *TR* in *LM*?" or "Is *TR* above *LM*?" or "Does *TR* go across *LM*?", etc. (cette problématique a également été abordée par Harris, 1990, et dans le L_0 -project de Feldman *et al.*, 1990; elle a été traitée plus complètement par Regier, 1995, au moyen de réseaux de neurones multi-couches traditionnels). Comme on l'a exposé plus haut, le schème sémantique impliqué dans la question aura pour effet de déclencher une chaîne particulière de transformations de l'image, suivie par un mécanisme de détection et de décision. Ces routines sont essentiellement des diffusions-propagations à partir des contours des objets.

L'application est une Applet Java. La figure 14 montre sa GUI (Graphical User Interface). La première fenêtre est celle de l'input. On peut soit y dessiner le *TR* et le *LM* (en utilisant les deux boutons de la souris) soit y importer une image et la segmenter en *TR* et *LM*. Le bouton "create image" passe à la seconde fenêtre l'input structuré en 3 couches: 1=*TR*, 2=*LM*, 3=*TR+LM* (cette structure correspondant à un format propre de l'Applet). Les autres fenêtres permettent d'enchaîner plusieurs transformations comportant chacune ses paramètres de contrôle (voir tableau 1). Elles comportent un bouton revert (!) inversant l'image, un bouton de seuillage (%) pour les niveaux de gris ou les couleurs RGB, un bouton Couleur/NB ainsi que deux ascenseurs permettant de contrôler les niveaux de couleurs ou de gris, et en particulier de définir des échelles *non monotones* permettant de faire *alterner* des niveaux clairs/sombres et, par là-même de visualiser des lignes de niveau.

Filtres	Paramètres
Copy	
Translation	<ul style="list-style-type: none"> • dx, dy (composantes); • <code>gray level</code>.
Heat (EDP de diffusion)	<ul style="list-style-type: none"> • x <code>coeff</code>, y <code>coeff</code> (coefficients suivant les 2 axes, x <code>coeff</code>=y <code>coeff</code> pour une diffusion isotrope); • <code># iterations</code> (0=indéfini); • <code>from 1 to 2</code> (contrôle de la diffusion du <i>TR</i> vers le <i>LM</i>).
Blum (EDP de propagation)	<ul style="list-style-type: none"> • <code>coeff</code> (coefficient de la propagation).
Skiz	
Gauss	<ul style="list-style-type: none"> • x <code>sigma</code>, y <code>sigma</code> (largeurs des gaussiennes dans les directions x, y); • <code>border gray</code> (niveau de gris du bord).
Contour	
Skeleton	<ul style="list-style-type: none"> • <code># iterations</code>.
Expansion	<ul style="list-style-type: none"> • <code># iterations</code>; • <code>horizontal/vertical</code>; <code>free/solid</code> (l'expansion solide s'arrête dès qu'elle rencontre un obstacle); • <code>1&2 / 1>2 / 1<2</code> (contrôle de l'expansion).
Morphology	<ul style="list-style-type: none"> • <code># neighbors</code> (# minimal de pixels voisins qui doivent être noirs pour que la transformation soit appliquée); • <code># iterations</code>; • <code>Dilation / Erosion / Closure / Opening</code>.

Tableau 1: Les paramètres de contrôle de l'Applet.

Enfin les boutons en bas à droite permettent de poser des questions à propos de l'applicabilité d'une préposition à l'input. Chaque question déclenche une séquence de traitements permettant d'arriver à un critère de décision catégoriel.



Figure 14: L'interface de l'Applet.

6.2. Exemple 1: “La balle dans la boîte”

Le premier exemple traite de la propriété d'inclusion. On considère tout d'abord une scène extrêmement simple, dans laquelle TR et LM sont deux objets fermés et TR est inclus dans LM . Par exemple, la figure 15 représente une balle dans une boîte.

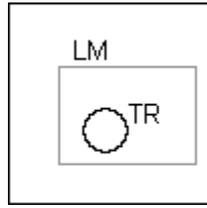


Fig. 15: Représentation schématique de “la balle (*TR*) dans la boîte (*LM*)”.

6.2.1. Diffusion : équation de la chaleur

Notre interprétation morphologique de “in” est que la question “*TR in LM ?*” consiste à déterminer si une *expansion de TR est bloquée ou non par LM*, c’est-à-dire si une propagation d’activité partant des contours de *TR* réussit ou non à parvenir sur les bords ∂W ¹⁰ de la fenêtre visuelle W ¹¹ en un temps $t \leq t_{max}$ où t_{max} est le temps de traversée de W .

Une première possibilité consiste à utiliser un processus de diffusion simple tel que celui de la chaleur. L’équation s’écrit $-\mu(\partial a/\partial t) = \Delta a$, où $a(x,y,t)$ est le taux d’activité au point (x,y) et au temps t , $\Delta a = \partial^2 a/\partial x^2 + \partial^2 a/\partial y^2$ le laplacien spatial de a , et $-\mu$ un coefficient faible.¹² Sans conditions aux limites particulières, ce processus relaxe vers un paysage uniforme $a = a_0$, où a_0 est une constante arbitraire. Ici, nous fixerons l’activité (par “clamping”) sur les contours de *TR* et *LM* à deux valeurs constantes différentes, soit, pour tout t : $a(\partial LM, t) = 0$ et $a(\partial TR, t) = 1$. Pour simplifier, l’état initial sur le reste du domaine est fixé à 0 partout ailleurs (bien que ceci n’ait pas d’influence sur l’état d’équilibre final). Métaphoriquement, *TR* joue le rôle d’une “source de chaleur” en expansion et *LM*, celui d’un “obstacle froid” tendant à maintenir des valeurs nulles autour de lui.

Ainsi, nous postulons que la *question “TR in LM ?” active les bords de TR et LM de façon spécifique*. Le résultat en est que le paysage d’activité relaxe vers un gradient entre *TR* et *LM* (décroissance de 1 vers 0), et, précisément parce que *TR* est entièrement inclus dans *LM*, l’activité au dehors de *LM* reste uniformément nulle (figure 16).

¹⁰ On note traditionnellement ∂W le bord d’un domaine W . Topologiquement, ∂W est la différence entre la fermeture et l’intérieur de W . Nous ne considérons que des cas simples où W est fermé et où ∂W est une courbe continue, différentiable par morceaux.

¹¹ La définition de la fenêtre visuelle dépend évidemment du contexte. Souvent, ce sera le champ visuel, mais elle pourra aussi être plus large. Nous n’entrerons pas ici dans ces difficultés.

¹² $\partial^n f/\partial x^n$ est la dérivée partielle d’ordre n de f par rapport à x . Δ est une dérivée seconde isotrope (i.e. invariante par rotation).

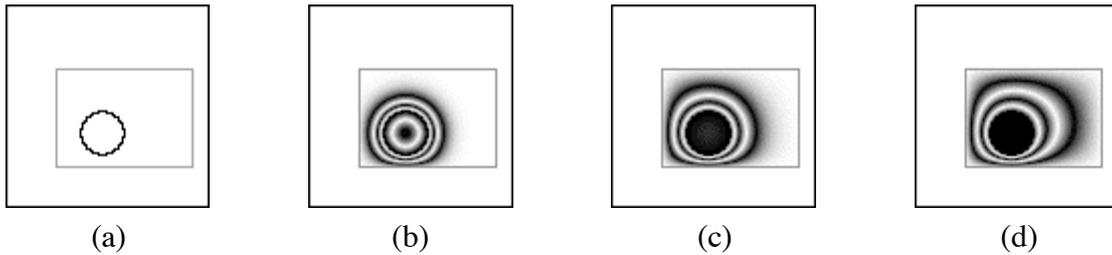


Figure 16: Processus de diffusion appliqué à “la balle dans la boîte”. (a) état initial $t = 0$; (b) état $t = 100$; (c) état $t = 300$; (d) état $t = 2000$. Avec des valeurs entre 0 et 1, le code de niveaux de gris est choisi pour faire apparaître les fronts de diffusion (valeurs 0 / 0.4 / 0.8 en blanc, valeurs 0.2 / 0.6 / 1 en noir). Le contour de TR (la balle) est fixé à 1, celui de LM (la boîte) à 0 (ici, exceptionnellement en gris pour la rendre visible).

Ainsi, le fait qu’aucune activité ne diffuse en dehors du domaine de LM pour atteindre le bord ∂W peut être pris comme un indice *catégoriel* caractéristique de l’inclusion de TR dans LM . Deux contre-exemples sont montrés dans la figure 17: dans tous les cas, l’activité sur les bords de l’image augmente rapidement vers des valeurs positives.



Figure 17: Processus de diffusion appliqué à deux contre-exemples de “la balle dans la boîte”. (a) état initial $t = 0$; (b) état $t = 500$; (c) état initial $t = 0$; (d) état $t = 300$. Le code de gris est le même que dans la figure précédente.

Insistons sur le fait que, *quelles que soient les formes métriques de TR et LM* , le processus d’analyse réduit l’applicabilité de “in” à un critère binaire purement *catégoriel* : il y a-t-il oui ou non de l’activité ayant rejoint le bord ∂W avant t_{max} ?

6.2.2. Transformations morphologiques

Nous avons noté au § 5.3 que les processus de diffusion gouvernés par des EDP sont computationnellement très lourds et très lents, et donc peu adaptés à des solutions pratiques. Il est difficile d’estimer la durée qu’ils requièrent pour atteindre un état d’équilibre approximatif considéré comme satisfaisant.

Nous avons vu également que des algorithmes beaucoup plus maniables sont fournis par la morphologie mathématique, avec, à la base, les deux routines complémentaires de dilatation et d'érosion. Dans la version la plus simple pour les images pixélisées binaires la règle de dilatation est la suivante : un pixel blanc devient noir si et seulement si au moins l'un de ses quatre plus proches voisins est déjà noir (inverser les termes noir et blanc pour l'érosion). Une itération consistant à visiter tous les pixels exactement une fois, le nombre d'itérations est donc à peu près égal à l'augmentation (la diminution) en épaisseur de la figure.

La figure 18 montre le processus de dilatation appliqué à la balle dans la boîte, où la boîte agit à nouveau comme un obstacle, les pixels qui la composent étant clampés dans l'état blanc (bien qu'ils soient rendus visibles en gris sur la figure).

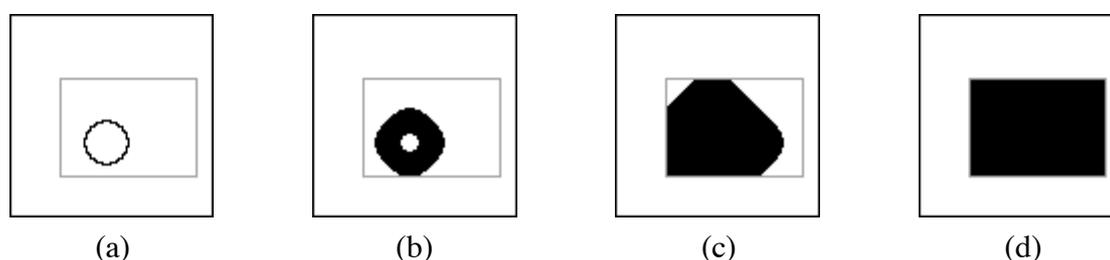


Figure 18: Routine de dilatation appliquée à “la balle dans la boîte”. (a) état initial $t = 0$; (b) itération $t = 6$; (c) itération $t = 24$; (d) arrêt à l'itération $t = 60$. Comparer à l'état $t = 2000$ de la figure 16: le gain opératoire est spectaculaire.

Puisque les pixels du contour de LM ne peuvent pas devenir noirs, l'expansion s'arrête au bout d'un petit nombre d'itérations. Ici aussi, TR est déclaré être inclus dans LM si et seulement si aucun pixel noir n'est détecté sur le bord de l'image (critère catégoriel binaire).

6.2.3. Zones d'influence, Skiz et cut-locus

Une alternative intéressante au scénario précédent consiste à faire jouer un rôle actif également au LM en provoquant une *contre-expansion* vers TR (comme précédemment par dilatation). La structure résultante est (cf. § 5.4) une “ligne de partage” ou “séparatrice” marquant la rencontre des deux zones d'influence des objets (figure 19).

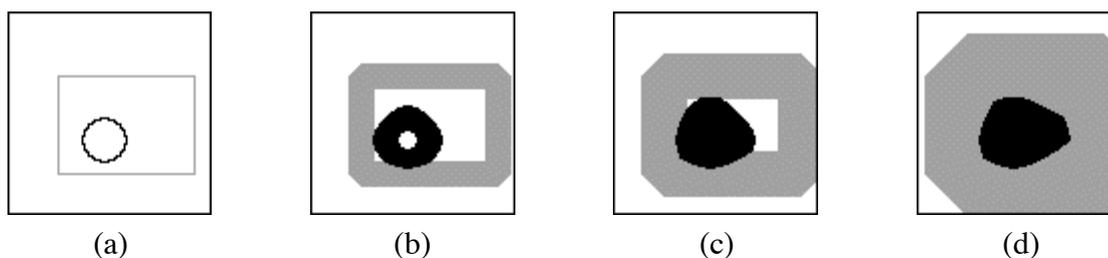


Figure 19: Dilatation simultanée de TR et LM , appliquée à “la balle dans la boîte”. (a) état initial $t = 0$; (b) itération $t = 6$; (c) itération $t = 11$; (d) arrêt de la croissance à l’itération $t = 21$. Comme dans les figures précédentes, TR est codé en noir et LM en gris.

Le critère d’inclusion doit ici distinguer entre deux types d’activités, celle de TR et celle de LM , et vérifier que la première n’est pas détectée sur les bords.

Notons que le contour du TR après expansion (la séparatrice) est une bonne approximation du *cut-locus externe*, qui est l’axe de symétrie généralisé du *fond* (la zone complémentaire entre la balle et la boîte). Ce cut-locus (ici, replié et formant une boucle) peut être également mis en évidence par la squelettisation du complémentaire (figure 20).

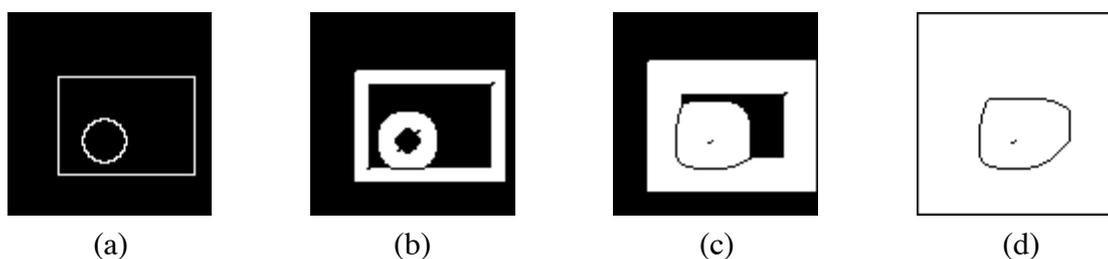


Figure 20: Squelettisation du fond extérieur à TR et LM , appliquée à “la balle dans la boîte”. (a) état initial $t = 0$; (b) itération $t = 6$; (c) itération $t = 16$; (d) arrêt de l’algorithme à l’itération $t = 64$. Noter également le cut-locus interne de la balle, réduit à un “point” (petit segment de 3 pixels).

Remarque

Techniquement parlant les algorithmes de squelettisation érodent progressivement le domaine noir, en préservant les configurations locales de segments d’épaisseur 1. A chacun des 256 voisinages possibles de 3×3 pixels autour d’un pixel noir, est associée une action: effacer ou non le pixel central. Le cut-locus idéal est constitué quant à lui de l’ensemble de tous les points équidistants de TR et LM (Blum, 1973). La figure 21 compare les lignes obtenues par double dilatation et par squelettisation avec la ligne d’équidistance.

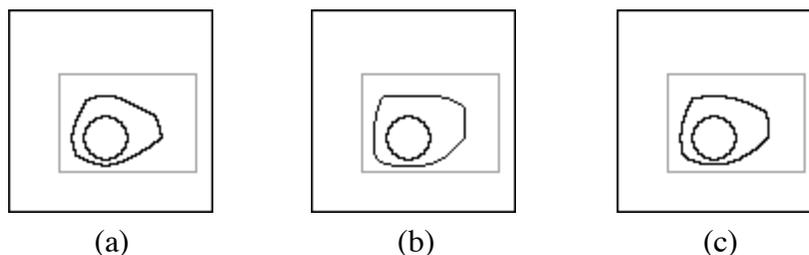


Figure 21: La séparatrice entre TR et LM calculée de trois façons différentes dans le cas de “la balle dans la boîte”. (a) approximation obtenue par double-dilatation; (b) approximation obtenue par squelettisation du complémentaire; (c) cut-locus exact calculé par équidistance.

Finalement, on adoptera de préférence la méthode de la double dilatation, plus rapide et moins *ad hoc* que les deux autres, et correspondant naturellement aux phénomènes d’expansion que l’on veut mettre ici en évidence. La figure 22 montre l’application de cette routine aux deux contre-exemples vus plus haut (figure 17).



Figure 22: Double dilatation appliquée à deux contre-exemples de “la balle dans la boîte”. L’activité de TR (en noir) a atteint les bords après, respectivement, 25 et 34 itérations.

6.3. Exemple 2: “L’oiseau dans la cage”

Considérons maintenant un cas d’inclusion plus complexe où le LM confinant le TR possède des ouvertures. Nous avons vu avec Talmy à la section 2 que la préposition “in” est neutre relativement à une telle modification. Dans la perspective de la sémantique active une telle neutralité n’est pas une donnée mais une construction, une propriété résultant d’un *prétraitement du LM*.

6.3.1. Prétraitement du LM

La difficulté du problème vient de ce que le domaine LM (“la cage”) présente un bord discontinu. Une solution consiste à opérer une *fermeture* préalable du LM avant

l'expansion (diffusion ou dilatation) du TR ("l'oiseau"). Une des façons les plus simples d'effectuer une telle fermeture sur un contour de forme totalement indéterminée est d'opérer une dilatation suivie non pas d'une érosion mais d'une squelettisation (figure 23).

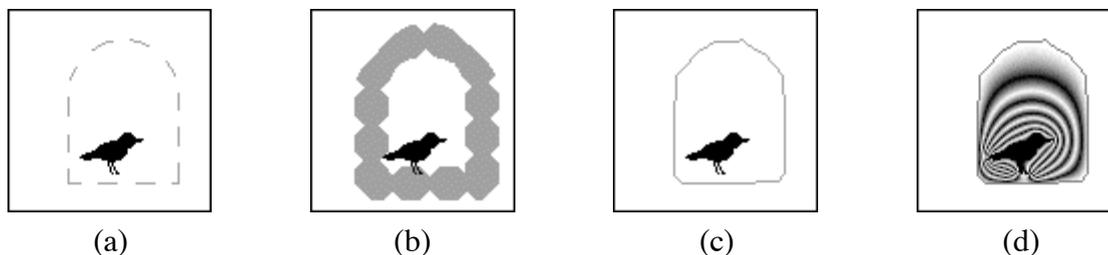


Figure 23: Détection de l'inclusion par fermeture préalable du LM , appliquée à "l'oiseau dans la cage". (a) état initial; (b) dilatation du LM (8 itérations); (c) squelettisation du LM (18 itérations); (d) expansion du TR (ici, par diffusion).

6.3.2. Traitement global par *Skiz*

Cependant, là aussi, il paraît plus simple de provoquer la double expansion simultanée de TR et LM faisant apparaître le cut-locus externe (figure 24).

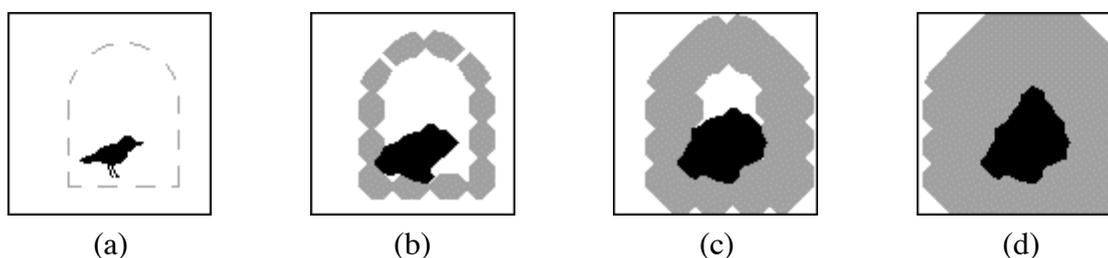


Figure 24: Détection de l'inclusion par double dilatation, appliquée à "l'oiseau dans la cage". (a) $t = 0$; (b) $t = 6$; (c) $t = 14$; (d) $t = 27$.

Remarquons aussi que, comme il est souhaitable, les détails morphologiques du TR n'ont que très peu d'influence sur le résultat de la dilatation. Ceux-ci sont effacés en un petit nombre d'itérations et, quelle que soit la forme d'origine, la forme dilatée s'approche rapidement d'un "blob" indifférencié.

6.4. Remarques

Pour développer une analyse morphologique d'une préposition apparemment aussi simple que "dans", il faut en fait des algorithmes beaucoup plus subtils et performants que ceux que nous venons d'esquisser.

Par exemple, pour tenir compte des cas où le LM ne possède pas seulement des ouvertures mais n'entoure pas complètement le TR , comme dans l'exemple "la pomme est dans la coupe", on peut utiliser une complétion sphérique LM^S du LM (une construction comme la sphère exinscrite) et une expansion "solide" du TR qui se bloque dès que le bord de diffusion atteint LM^S .

Dans les cas plus complexes comme par exemple "la cuillère est dans la tasse de café" (où seule la partie creuse de la cuillère est dans la tasse et où le manche est à l'extérieur) on doit faire intervenir des pré-traitements cognitifs de haut niveau, en l'occurrence prendre en compte la fonction instrumentale de la cuillère et sa décomposition méréologique canonique en "partie creuse" et "manche". La partie creuse étant fonctionnellement dominante, c'est par métonymie que porte préférentiellement sur elle l'application de la préposition "dans".

6.5. Exemple 3: "La lampe au-dessus de la table"

Considérons maintenant un exemple très différent, celui de la préposition "au-dessus" (above).

6.5.1. Prétraitement du LM

Le localisateur "au-dessus" fait fondamentalement intervenir le gradient gravitationnel haut/bas. Sa sémantique est non seulement géométrique (positionnelle) mais en quelque sorte "physique" (en un sens qualitatif). Dans un premier temps, l'analyse morphologique du schème de "supériorité" spatiale pourra donc utiliser une *expansion anisotrope* du TR selon l'axe vertical et une détection d'activité sur le bord *inférieur* du cadre ∂W de l'image. Si le TR se trouve "au-dessus" du LM , le LM fera obstacle à cette expansion et aucune activité ne parviendra en bas (cette idée a également été développée par Regier, 1988). Cependant, pour traiter correctement les cas possibles de décalage, une expansion horizontale préalable du LM est en général nécessaire (figure 25). Cela correspond bien au contenu sémantique de "au-dessus" qui recouvre en fait deux intuitions:

- (i) une intuition positionnelle: si TR tombait il serait arrêté par LM ;
- (ii) une intuition métrique: la côte du bas de TR est supérieure à celle du haut de LM .

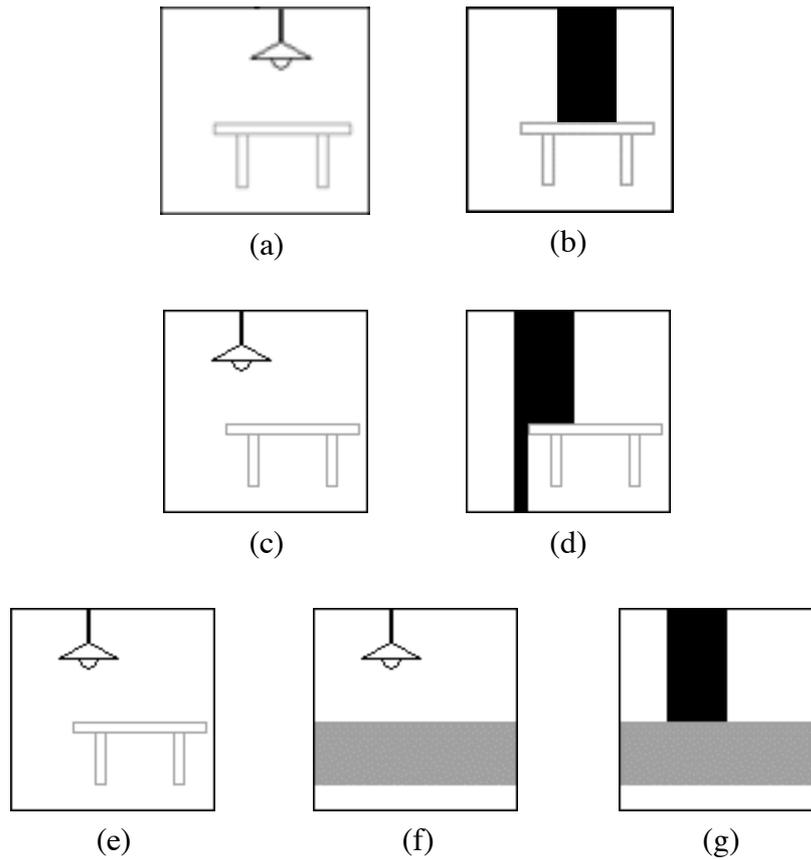


Figure 25: Détection de la supériorité par expansions successives appliquée à “la lampe au-dessus de la table”. (a), (b) *TR* au-dessus de *LM* prototypique; (c), (d) problème du décalage; (e), (f), (g) résolution du problème du décalage par expansion horizontale préalable du *LM*.

6.5.2. Traitement global par *Skiz*

A nouveau, l’expansion simultanée et isotrope des deux participants *TR* et *LM* semble fournir une bonne solution à toutes les situations. En effet, encore une fois, les zones d’influence et leur ligne de partage *Z* sont peu sensibles aux variations de la configuration initiale. Le critère catégoriel (binaire) de décision dépendra de l’intersection de la séparatrice *Z* avec le bord ∂W de la fenêtre visuelle *W*: *Z* ne doit pas intersecter la limite inférieure de *W* (figure 26).

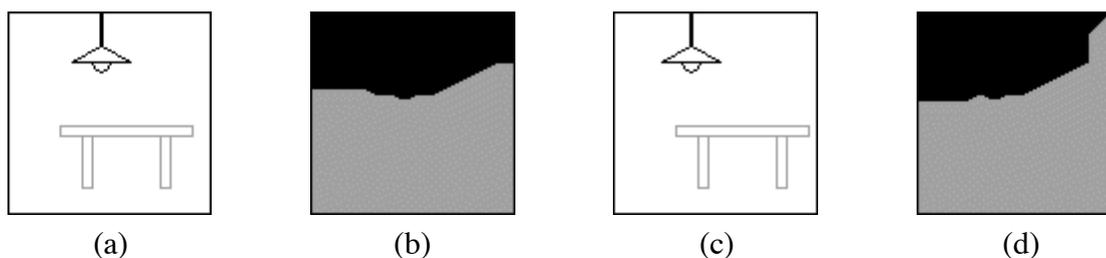


Figure 26: Détection de la supériorité par double dilatation, appliquée à “la lampe au-dessus de la table”. (a), (b) situation standard; (c), (d) situation décalée.

Notons que, de façon générale, l’intersection de Z avec ∂W fournit des informations essentielles sur la disposition relative de TR et LM . Par exemple, dans le cas “au-dessus”, elle permet de distinguer entre des configurations a priori semblables que certaines langues conceptualisent cependant par des éléments grammaticaux différents. La figure 27 en donne un exemple (“au-dessus” / “par-dessus”).

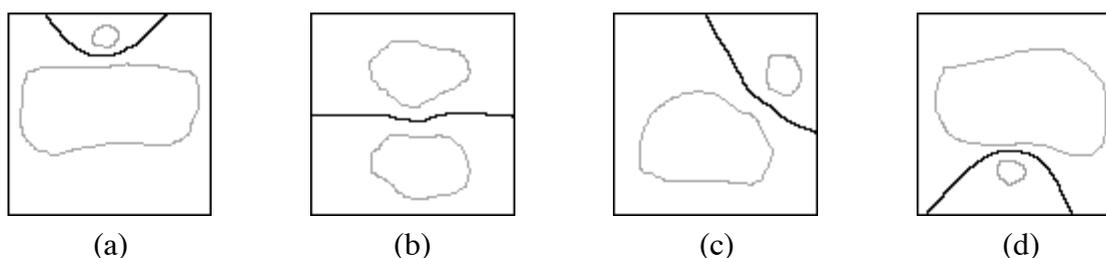


Figure 27: Différents cas de “ TR au-dessus de LM ”, distingués par la forme de la séparatrice Z et la position de ses deux points d’intersection avec le bord ∂W . Les deux blobs (contours gris) représentent TR et LM (du haut vers le bas, respectivement). La séparatrice Z (en noir) a été calculée exactement par équidistance (cf. figure 21-c). (a) petit TR au-dessus d’un grand LM : Z présente une courbure positive de type parabololoïde; (b) TR et LM de tailles semblables: Z est approximativement plate; (c) TR au-dessus et à côté de LM : Z est inclinée (type hyperboloïde) et intersecte l’un des bords verticaux du cadre; (d) grand TR au-dessus d’un petit LM : Z a une courbure négative (par exemple, en français, cette dernière situation sera plus souvent décrite avec le localisateur “par-dessus” que avec “au-dessus”: “le ciel est par-dessus le toit”).

6.5.3. Résolution d’un problème du L_0 -project

Ce qui précède fournit un excellent exemple de la façon dont des relations spatiales pouvant se déformer de façon *continue* peuvent être catégorisées de façon *discrète*. Cela résout l’un des problèmes clef du L_0 -project de Berkeley évoqué au § 1.1.

Considérons un TR = “disque” au-dessus d’un LM = “triangle” et déplaçons horizontalement TR . On passe continûment d’une situation “ TR au-dessus de LM ” à une

situation “ TR à côté de LM ”. Il faut par conséquent extraire de l’analyse morphologique un critère catégoriel. En considérant le Skiz de la configuration $TR-LM$ on peut facilement en définir un. Dans le cas “au-dessus” Z intersecte ∂W uniquement sur les côtés H-H (ou éventuellement G-H, H-D, G-D), alors que dans le cas “à côté” Z intersecte ∂W uniquement sur les côtés H-B (ou éventuellement G-G, ou D-D). La bifurcation se produit donc au passage de H-D (ou H-G) à H-B (figure 28).¹³

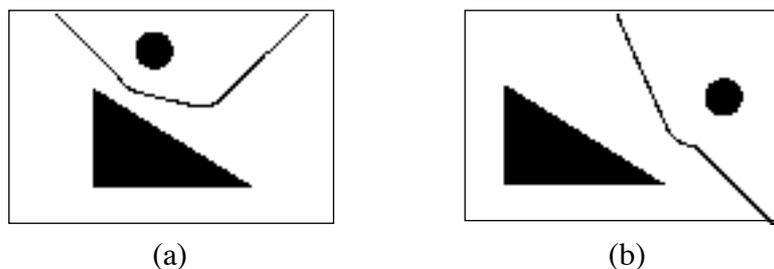


Figure 28: Passage continu de “au-dessus” (Z intersecte ∂W en H-H) à “à côté” (Z intersecte ∂W en H-B). La bifurcation se produit lorsque le point d’intersection en haut à gauche, après avoir descendu le côté G passe du côté B.

6.6. Exemple 4: “Zigzaguer à travers la forêt”

Nous prendrons comme dernier exemple celui de “across”, déjà abordé plus haut à la section 4. Dans le cas du concept de “transversalité” l’analyse morphologique est notablement différente. Comme nous l’avons vu, il s’agit essentiellement de “tubifier” la trajectoire du TR par une dilatation suivie d’une squelettisation et, d’autre part, d’extraire le cut-locus du LM , quel que soit le degré de complexité de sa texture. Ceci peut être également réalisé par une dilatation suivie d’une squelettisation. Il reste alors à détecter non pas une activité sur les bords, mais la *présence d’un point quadruple hétérogène*, signalant quelque part dans l’image un croisement entre deux lignes *d’origines différentes*¹⁴ (TR et LM) (figure 28).

¹³ Dans le cas où ∂W est circulaire (ce qui correspond mieux au champ visuel), on peut le découper en H/D/B/G en considérant le carré inscrit. On sait d’ailleurs qu’il existe dans le champ visuel des coordonnées canoniques (H/B, D/G) liées au corps propre.

¹⁴ Cette condition est très importante. Il faut soigneusement distinguer les points quadruples d’intersection entre deux branches respectives des cut-locus $Sk(TR)$ et $Sk(LM)$ et les points multiples que $Sk(TR)$ et $Sk(LM)$ peuvent présenter (génériquement des points triples).

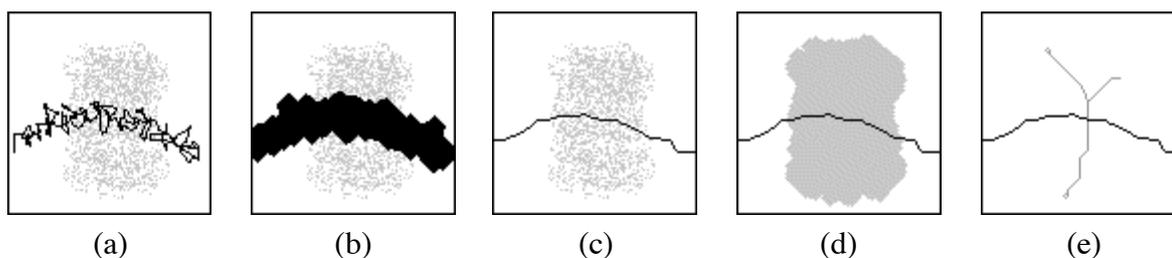


Figure 28: Le schème de transversalité. (a) L'input; (b) et (c) prétraitement du *TR* (dilatation+squelétisation); (d) et (e) prétraitement du *LM* (dilatation+squelétisation). On remarque dans (e) le crossing (point quadruple) entre $Sk(TR)$ et $Sk(LM)$ qui est caractéristique de "across".

On notera l'extraordinaire simplification des données apportée par la schématisation morphologique et la manière dont elle permet d'extraire l'invariant de transversalité. Les algorithmes morphologiques mettent en évidence le pattern de croisement \times . Celui-ci est ensuite détecté (on sait que de tels détecteurs existent dans le cortex visuel) et sa détection fournit le critère catégoriel binaire cherché. Notre hypothèse est que la *prédication* "*TR* traverse *LM*" ne devient possible que sur la base de cette schématisation morphologique préalable.

7. UN POSSIBLE SUBSTRAT NEURONAL: CORRÉLATIONS ET ONDES D'ACTIVITÉ

Dans cette dernière section, nous nous proposons d'esquisser très rapidement un lien entre le modèle dynamique d'inspiration physique présenté ci-dessus et une certaine famille de modèles dynamiques d'inspiration neurobiologique ou connexionniste qui accordent une place centrale aux *corrélations temporelles* entre signaux neuronaux. Pour cela, nous nous appuyons sur une analogie frappante entre les phénomènes d'"expansion" (diffusion, propagation), tels que nous venons de les schématiser sur un substrat pseudo-physique de pixels, et des phénomènes de *propagation d'ondes* d'activité qui peuvent émerger sur un réseau d'unités de type neuronal, c'est à dire des unités ayant des propriétés excitatrices semblables à celles des neurones (figure 29).

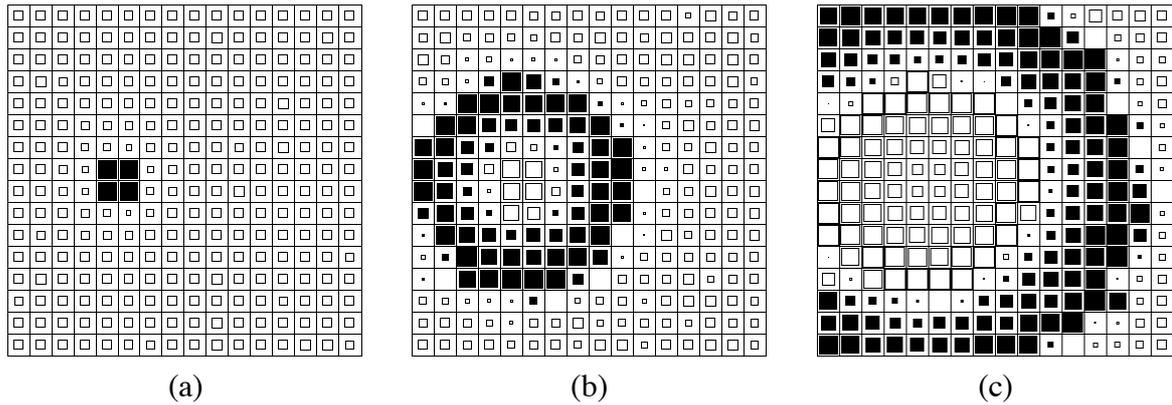


Figure 29: Émergence et propagation d'une onde d'activité sur un réseau de 16×16 unités neuronales “Bonhoeffer-van-der-Pol” faiblement couplées (voir plus bas).

L'activité de chaque unité est codée par un carré de taille proportionnelle à sa valeur absolue et de couleur noire/blanche correspondant à son signe. (a) Conditions initiales: l'impulsion est donnée par un petit groupe de 4 unités; (b) et (c) la dépolarisation se propage de voisin en voisin et fait apparaître un front d'onde circulaire.

Il peut s'agir, dans le cas classique, d'oscillateurs faiblement couplés, ou d'unités quasi oscillatoires (stochastiques ou non), ou encore, de façon très générale, d'éléments excitateurs sans propriétés particulières de périodicité (les fronts d'ondes dépendent alors d'une impulsion initiale). Les liens du réseau modélisant les connexions synaptiques suivront généralement un maillage topologique à peu près régulier (connectivité locale aux plus proches voisins dans une structure quasi plane, ou bien structure linéaire en “chaîne”: voir Abeles, 1982; Bienenstock, 1995) et posséderont éventuellement des délais de transmission non-nuls. L'activité globale du réseau s'organise alors de façon très différente selon tous ces paramètres. Parmi les nombreux modes de fonctionnement possibles, et sans vouloir entrer dans une analyse plus détaillée, on retiendra

- (a) les *synchronisations* de groupes,
- (b) les *propagations* d'ondes.

Les phénomènes de type (a) ont été à la base de la plupart des modèles connexionnistes de segmentation ou “binding” (voir par exemple, pour les débuts, Malsburg et Schneider, 1986; Gray et Singer, 1989; König et Schillen, 1991; Neven et Aertsen, 1992). Ces modèles reposent sur le principe du “temporal tagging” (étiquetage temporel) des domaines: toutes les unités d'un bloc connexe se mettent à osciller en phase et leur phase commune ϕ_1 est distincte de la phase commune ϕ_2 d'un autre bloc. Comme nous l'avons mentionné au début de la Section 6, ils offrent ainsi une solution plausible à la première étape d'*individuation* des participants *TR* et *LM* d'une scène visuelle.

Nous nous intéressons quant à nous aux phénomènes de type (b), particulièrement appropriés à l'étape suivante d'*interaction* entre *TR* et *LM*, à base d'expansions et

d'obstacles. En modifiant par exemple les délais de transmission des connexions, ou simplement les conditions initiales, on peut facilement sortir du mode (a) et créer des propagations d'activité prenant des formes très diverses : ondes planes, ondes circulaires, vortex (spirales), etc. Ces motifs plus ou moins complexes sont en général décrits par un *champ de phases* non-uniforme $\{\phi_i, i = 1, \dots, N\}$, N étant le nombre d'unités (voir Kuramoto, 1984; Kuramoto et Nishikawa 1987). Par exemple, la figure 19 montre l'apparition d'un front d'onde sur un réseau de 16×16 unités neuronales stochastiques de type Bonhoeffer-van-der-Pol (appelées "stochastic excitable elements" chez Kurrer *et al.*, 1990), c'est à dire obéissant à un système d'équations différentielles simulant la génération de potentiels d'action ("spikes") sur l'axone. Chaque neurone est connecté à ses 4 plus proches voisins à travers un coefficient de transmission synaptique relativement faible (ce couplage se traduisant par un terme supplémentaire dans les équations de fonctionnement). Une impulsion initiale donnée au centre du réseau provoque alors des émissions de spikes de proche en proche. Finalement, pour modéliser l'effet d'un obstacle, on pourra jouer sur la force du couplage: la propagation peut être arrêtée par des coupures de connexions locales (coefficient de transmission proche de 0).

En résumé, au-delà de ces nombreuses variantes, nous espérons avoir fait entrevoir les riches possibilités d'implémentation, au moyen de mécanismes neuronaux plausibles, des processus d'expansions et d'obstacles à l'oeuvre dans la sémantique morphologique active, suggérant aussi sa possible présence dans les aires corticales visuelles et associatives.

Cette plausibilité et cette possibilité sont réelles car les spécialistes de neurosciences cognitives intégrées sont de plus en plus nombreux à considérer qu'il existe un neuro-calcul très général fondé sur des synchronisations et des propagations d'activités. Par exemple, dans un remarquable target article de *Behavioral and Brain Sciences*, W.A. Philipps et W. Singer ont bien mis en évidence ce qu'ils appellent les "common foundations for cortical computation".¹⁵ Le cortex est un isocortex. Son mode de traitement de l'information est très homogène. L'idée de base est que ses unités sont des "local processors with contextual guidance" qui possèdent deux types d'inputs:

- (i) leurs *champs récepteurs* qui agissent comme des filtres adaptatifs sélectivement sensibles à certains traits des stimuli et prennent en charge la spécialisation fonctionnelle du traitement;
 - (ii) leurs *champs contextuels* qui les relient à l'activité des processeurs voisins et prennent en charge l'intégration contextuelle au moyen de synchronisations temporelles.
- La synchronisation contextuelle joue un rôle crucial dans l'action computationnelle des processeurs car elle renforce de façon sélective, sans pour autant la dégrader,

¹⁵ Philipps, Singer [1996].

l'information issue des champs récepteurs qui est la plus cohérente relativement au contexte.

Dans les modèles présentés ici ces principes sont mis en oeuvre en considérant que l'information fournie par les champs récepteurs (ici réduite à une intensité de pixel) déclenche des ondes d'activités synchronisées et que, dans la mesure où il y a en général plusieurs sources, ces ondes entrent en conflit et engendrent des structures virtuelles organisatrices.

8. CONCLUSION

En utilisant des algorithmes morphologiques de nature géométrico-physique nous avons tenté de montrer sur un exemple élémentaire que les grammaires cognitives sont en principe mathématisables au même titre que les grammaires formelles, mais avec des outils tout différents. Nous avons montré comment les catégorisations discrètes de formes continûment déformables pouvaient être généralisées de l'analyse des objets à celle de configurations d'objets (de scènes) et comment, ce faisant, on pouvait implémenter la sémantique de relations spatiales statiques d'une façon qui soit compatible avec l'infinité continue de leurs référents possibles. Nous avons donc construit — et de façon algorithmiquement effective — ce que l'on pourrait appeler des *schèmes prépositionnels*.

Nous avons également vu qu'une telle sémantique active est assez subtile et, dès les niveaux apparemment les plus triviaux, fait apparaître dans toute sa complexité le problème de la catégorisation conceptuelle. Comme nous aimons à le dire, "tout concept est un algorithme qui s'ignore" et son applicabilité à ses référents suppose que l'on ait réussi à résoudre le problème du schème associé. Tout concept à contenu empirique (i.e. possédant des référents sensibles) est un schème et la linguistique repose donc sur l'explicitation algorithmique de ce que Kant appelait "cet art caché dans les profondeurs de l'âme". La conséquence en est qu'il existe un principe de parallélisme entre la diversité d'éléments sémantiques de base comme la classe fermée des prépositions et celle des algorithmes qui les rendent applicables et réalisent la compétence sémantique associée. Nous avons posé le problème de cette compétence. Nous pensons l'avoir résolu en ayant montré que la diversité des algorithmes dérive de principes généraux et universels de traitement d'information et de calcul neuronal.

Ceci dit, il ne s'agit dans cet article que d'une composante extrêmement limitée d'un programme de recherche beaucoup plus vaste. Un nombre considérable de problèmes difficiles n'ont pas été abordés. Citons-en quelques uns.

(i) Celui du passage des relations statiques aux relations dynamiques. Il est traité entre autres dans Petitot [1989b], [1991b].

- (ii) Celui des usages abstraits et figuratifs des schèmes prépositionnels. Nous pensons que c'est du côté des théories cognitives de la métaphore et du blending (cf. § 1.1) qu'il faut chercher la réponse.
- (iii) Celui de l'apprentissage. Certains éléments de réponse se trouvent dans Regier [1995].
- (iv) Le passage à la *syntaxe*. Sur ce point on pourra se référer aux travaux de l'un des auteurs (JP) remontant aux propositions de René Thom sur une syntaxe topologique d'attracteurs. A la fin des années 60, R. Thom et C. Zeeman ont été les premiers à introduire l'idée qu'un contenu mental était descriptible comme un attracteur d'une dynamique neurale sous-jacente "rapide", le flux temporel de ces contenus comme une dynamique "lente" opérant sur leurs dynamiques "rapides" et les événements mentaux comme des bifurcations d'attracteurs. Qui plus est, en introduisant des paramètres d'ordre (activités moyennes), ils montrèrent que l'on pouvait réduire drastiquement le nombre de degrés de liberté des systèmes considérés et passer ainsi du niveau neuronal "micro" au niveau psychologique "macro".¹⁶ Cette approche a été appliquée à la syntaxe en passant, comme ici, par l'analyse de la perception. L'idée de base consistait à utiliser ce que l'on appelle en linguistique *l'hypothèse localiste* qui fonde la sémantique casuelle des langues (i.e. la sémantique actantielle des verbes) dans la catégorisation des relations spatio-temporelles (cf. Petitot [1989a], [1991b], [1993], [1994], [1995]). Par métaphore et blending on peut alors comprendre comment ces schèmes verbaux de relations actantielles peuvent s'appliquer de façon abstraite et figurative à des situations extrêmement variées.

¹⁶ Trop révolutionnaires pour la fin des années 60 ces idées ont été largement incomprises. Les modèles connexionnistes qui ont depuis déferlé sur les sciences cognitives en constituent une implémentation dans des réseaux de neurones explicites. La justice la plus élémentaire commande donc de considérer R. Thom et C. Zeeman comme des précurseurs visionnaires dont la priorité scientifique est incontestable.

BIBLIOGRAPHIE

- Abeles, M. 1982. *Local Cortical Circuits: An Electrophysiological Study*. Springer-Verlag, Berlin.
- Bienenstock, E. 1995. A model of neocortex. *Network: Computation in Neural Systems* **6**, 179-224.
- Blum, H. 1973. Biological shape and visual science. *Journal of Theoretical Biology* **38**, 205-287.
- Brooks, R. A. 1986. A robust layered control system for a mobile robot. *IEEE Journal of Robotics and Automation*, **RA-2**(1), 14-23.
- Feldman, J. A., Lakoff, G., Stolcke, A., and Hollbach Weber, S. 1990. Miniature language acquisition: A touchstone for cognitive science. In *Proceedings of the 12th Annual Conference of the Cognitive Science Society*, pp. 686–693. MIT Press, Cambridge, Mass.
- Fillmore, C. J. 1977. The case for case reopened. In *Grammatical Relations*, P. Cole & J. M. Sadock, eds., pp. 59-82. Academic Press, New York.
- Gibson, J. J. 1979. *The Ecological Approach to Visual Perception*. Houghton-Mifflin, Boston.
- Gray, C. M., and Singer, W. 1989. Stimulus-specific neuronal oscillations in orientation columns of cat visual cortex. *Proc. Natl. Acad. Sci. USA* **86**, 1698-1702.
- Grush, R., Mandelblit, N.
- Hamy, H. 1997. *Méthodes géométriques multi-échelle en vision computationnelle*, Thèse, Paris, Ecole Polytechnique.
- Harris, C. 1990. Connectionism and cognitive linguistics. *Connection Science* **2**(1/2), 7-33.
- Jackendoff, R. 1983. *Semantics and Cognition*. MIT Press, Cambridge, Mass.
- König, P., and Schillen, T. B. 1991. Stimulus-dependent assembly formation of oscillatory responses: I. Synchronization. *Neural Comp.* **3**, 155-166.
- Kuramoto, Y. 1984. *Chemical Oscillations, Waves, and Turbulence*. Springer-Verlag, Berlin.
- Kuramoto, Y., and Nishikawa, I. 1987. Statistical macrodynamics of large dynamical systems. Case of a Phase transition in oscillator communities. *J. Stat. Phys.* **49**(3/4), 569-605.
- Kurrer, C., Nieswand, B., and Schulten, K. 1990. A model for synchronous activity in the visual cortex. In *Self-Organization, Emerging Properties and Learning*, A. Babloyantz, ed. Plenum Press.
- Lakoff, G. 1987. *Women, Fire, and Dangerous Things: What Categories Reveal About the Mind*. University of Chicago Press, Chicago.

- Lakoff, G. 1988. A suggestion for a linguistics with connectionist foundations. In *Proceedings of the 1988 Connectionist Models Summer School*. M. Kaufman.
- Langacker, R. W. 1987. *Foundations of Cognitive Grammar. Volume I: Theoretical Prerequisites*. Stanford University Press, Stanford.
- Mallat, S. 1989. "Multifrequency Channel Decompositions of Images and Wavelet Models", *IEEE Transactions on Acoustics, Speech, and Signal Processing* **37** (12) : 2091-2110.
- Marr, D. 1982. *Vision*. Freeman, San Francisco.
- Morel, J.-M., and Solimini, S. 1995. *Variational Methods in Image Segmentation*. Birkhauser, Boston.
- Mumford, D. 1993. Pattern-theory: A unifying perspective. In *Proceedings of the 1st European Congress of Mathematics*. Birkhauser, Boston.
- Neven, H., and Aertsen, A. 1992. Rate coherence and event coherence in the visual cortex: A neuronal model of object recognition. *Biol. Cybern.* **67**, 309-322.
- Perona, P., and Malik, J. 1990. Scale-space and edge detection using anisotropic diffusion. *IEEE Trans. PAMI* **12**(7), 629-639.
- Petitot, J. 1989a. "Hypothèse localiste, Modèles morphodynamiques et Théories cognitives : Remarques sur une note de 1975", *Semiotica* **77**(1/3), 65-119.
- Petitot, J. 1989b. "Modèles morphodynamiques pour la Grammaire cognitive et la Sémiotique modale", *Recherches Sémiotiques/ Semiotic Inquiry*, **9**(1/2/3), 17-51.
- Petitot, J. 1991a. "Why Connectionism is such a Good Thing. A Criticism of Fodor's and Pylyshyn's Criticism of Smolensky", *Philosophica*, **47**(1), 49-79.
- Petitot, J. 1991b. Syntaxe topologique et grammaire cognitive. *Langages* **103**, 97-128.
- Petitot, J. 1992. *Physique du Sens*. Editions du CNRS, Paris.
- Petitot, J. 1993. "Natural Dynamical Models for Cognitive Grammars", *Language and Cognition* (F. Manjali ed.), 81-104, New Dehli, Bahri Publications.
- Petitot, J. 1994. Attractor syntax: Morphodynamics and cognitive grammars. In *Continuity in Linguistic Semantics*, C. Fuchs & B. Victorri, eds., pp. 167-187. John Benjamins, Amsterdam.
- Petitot, J. 1995. Morphodynamics and attractor syntax: Dynamical and morphological models for constituency in visual perception and cognitive grammar. In *Mind as Motion*, T. van Gelder & R. Port, eds., pp. 227-281. MIT Press, Cambridge, Mass.
- Philipps, W.A., Singer, W. 1996. In Search of Common Foundations for Cortical Computation, *Brain and Behavioral Sciences*, <ftp://ftp.princeton.edu/pub/harnad/BBS/.WWW/bbs.philipps.html>.
- Regier, T. 1988. Recognizing image-schemas using programmable networks. In *Proceedings of the 1988 Connectionist Models Summer School*. M. Kaufman.

- Regier, T. 1995. A model of the human capacity for categorizing spatial relations. *Cognitive Linguistics* 6(1), 63–88.
- Schmitt, M., Mattioli, J., 1994. *Morphologie mathématique*, Paris, Masson.
- Serra, J., 1982. *Image Analysis and Mathematical Morphology*, London, Academic Press.
- Steels, L. 1994. Building agents with autonomous behavior systems. In *The 'Artificial Life' Route to 'Artificial Intelligence': Building Situated Embodied Agents*, L. Steels & R. Brooks, eds. Lawrence Erlbaum Associates, New Haven.
- Sweetser, E. 1990. *From Etymology to Pragmatics: Metaphorical and Cultural Aspects of Semantic Structure*. Cambridge University Press.
- Talmy, L. 1978/1987. The relation of grammar to cognition. In *Topics in Cognitive Linguistics*, B. Rudzka-Ostyn, ed. John Benjamins, Amsterdam.
- Talmy, L. 1983. How language structures space. In *Spatial Orientation: Theory, Research, and Application*, H. Pick and L. Acredolo, eds., pp. 225-282. Plenum Press, New York.
- Talmy, L. 1985. Force Dynamics in Language and Thought. In *Parasession on Causatives and Agentivity, Papers from the 21st Regional Meeting of the Chicago Linguistic Society*, pp. 293-337.
- Thom, R. 1980. *Modèles Mathématiques de la Morphogénèse*. Christian Bourgois, Paris.
- van Gelder, T., and Port, R., eds. 1995. *Mind as Motion*. MIT Press, Cambridge, Mass.
- von der Malsburg, C. 1981. *The correlation theory of brain function*. Internal report 81-2, Max-Planck Institute for Biophysical Chemistry, Dept. of Neurobiology, Göttingen, Germany.
- von der Malsburg, C., and Bienenstock, E. 1986. Statistical coding and short-term synaptic plasticity: A scheme for knowledge representation in the brain. In *Disordered Systems and Biological Organization*, E. Bienenstock, F. Fogelman and G. Weisbuch, eds., pp. 247-272. Springer-Verlag, Berlin.
- Whitaker, R. T. 1993. Geometry-limited diffusion in the characterization of geometric patches in images. *CVGIP: Image Understanding* 57(1), 111–120.